TECHNISCHE UNIVERSITÄT MÜNCHEN

Institut für Photogrammetrie und Kartographie
Lehrstuhl für Methodik der Fernerkundung

Automatic Image to Image Registration for Multimodal Remote Sensing Images

Sahil Suri

Vollständiger Abdruck
der von der Fakultät für Bauingenieur- und Vermessungswesen
der Technischen Universität München
zur Erlangung des akademischen Grades
eines Doktor-Ingenieurs (Dr.-Ing.)
genehmigten Dissertation**.**

Vorsitzender**:** Univ.-Prof. Dr.-Ing. Uwe Stilla

Prüfer der Dissertation: 1. Univ.-Prof. Dr.-Ing. habil. Richard Bamler
2. Univ.-Prof. Dr.-Ing. Stefan Hinz,
Universität Karlsruhe (TH)

Die Dissertation wurde am 20.11.2009 bei der Technischen Universität München eingereicht und durch die Fakultät für Bauingenieur- und Vermessungswesen am 02.02.2010 angenommen.

Dedicated to my Parents

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

This dissertation entitled "Automatic Image to Image Registration for Multimodal Remote Sensing Images" would not have been completed without imperative support and contribution from a lot of people besides me.

First of all I would like to express my gratitude towards Prof. Dr. Richard Bamler, director at the *Remote Sensing Technology Institute (IMF)* at the *German Aerospace Centre (DLR)* for giving me an opportunity and periodical counsel to work on this very interesting topic of research.

I would like to thank all my colleagues at *Photogrammetry and Image Analysis (PB)* group for giving me suggestions at various stages of my research and as well as for fostering a constructive atmosphere inculcating novel ideas and their implementation, that were very much required for this dissertation. Specifically, I would like to thank Mr. Manfred Lehner for introducing me to our image processing system (Xdibias) and giving me useful ideas during very early stages of this work. I thank Dr. Danielle Hoja, my project coordinator, for arranging the datasets and shielding me from all bureaucratic affairs during the course of my research. I need to mention contribution from Mr. Mathias Schneider in helping me to make my Xdibias module CAMRI (Completely Automated Mutually Registered Images) bug free and functional for different applications in our institute. I sincerely thank Mr. Rolf Stätter for carefully reading and correcting errors to bring my dissertation in this presentable form. I specially appreciate Mr. Peter Schwind for a very significant contribution in developing new ideas and results, presented in this dissertation. Time spent with him discussing problems and their possible solution always gave me an important second perspective. I also extend my word of thanks to Mr. Johannes Uhl and Mr. Sebastian Türmer, for working directly with me at different stages of this work.

This research would not have completed without some quality mentoring from Dr. Peter Reinartz, head of our research group. I would always cherish this three year period that I spent working under his guidance with a degree of freedom and flexibility that any Ph.D. candidate desires for. I also need to thank Prof. Dr. Stefan Hinz, *Institut für Photogrammetrie und Fernerkundung, Universität Karlsruhe,* for invaluable advice at different stages of this research. I always looked

forward to my interactions with him at various summer schools, workshops and project meetings for important comments and suggestions. I also thank him for being one of my supervisors for submitting this dissertation at *Technische Universität München (TUM)*. I want to acknowledge another special mentor, Prof. Dr. Manoj Arora, *Geomatics Engineering Section, Department of Civil Engineering, Indian Institute of Technology (IIT) Roorkee*, though he was physically never present during this research but this work would not have been possible without immense faith and confidence that he bestowed over me during my stay at IIT Roorkee.

Finally, I express my deepest gratitude to all my family, friends, well wishers and especially Dr. Madhuri Shivalkar in helping me to achieve that I always thought to be difficult and beyond my reach.

ABSTRACT

During the last decades, remote sensing sensors have undergone a rapid development in terms of both data quantity and characteristics. With advancements in remote sensing technology, the use of satellite images in disparate fields has received a tremendous boost. Few of these include generation of 3D models and topographic maps, early warning systems, urban growth, damage assessment, crisis information management and disaster mitigation. These applications normally utilize image processing techniques like image fusion, change detection, GIS overlay operations or 3D visualization which requires registered images procured from different sources.

Image registration is a fundamental task in remote sensing image processing that is used to match two or more images taken, for example, at different times, from different sensors or from different view points. A lot of automation has been achieved in this field but ever sprouting data quality and characteristics compel innovators to design new and/or improve existing registration techniques. In literature, image registration methodologies are broadly classified into intensity and feature based approaches. In this dissertation, we have evolved and combined two distinct techniques from each of the broad classes to extend their applicability for answering contemporary challenges in remote sensing image registration.

Generally, remote sensing applications need to accommodate images from different sensors/modalities; reason might be specific application demands or data availability. For example in case of a natural calamity, decision makers might be forced to use old archived optical data with a newly acquired (post-disaster) SAR image. Misalignment within procured SAR and optical imagery (both orthorectified) in such scenarios is a common phenomenon and these registration differences need to be taken care of prior to their joint application. Considering the recently available very high resolution (VHR) data available from satellites like TerraSAR-X, Risat, IKONOS, Quickbird, ALOS etc, registering these images manually is a mammoth task (due to volume and scene characteristics). Intensity based similarity metrics like mutual information (MI) and cluster reward algorithm (CRA) have been found useful for achieving registration of SAR-optical data from satellites like Landsat, Radarsat, SPOT, and IRS but still their application for high resolution data especially acquired over urban areas is limited. In this dissertation, we analyze in detail the performance of MI for very high resolution remote sensing images and

evaluate (feature extraction, classification, segmentation, discrete optimization) for improving its accuracy, applicability and processing time for VHR images (mainly TerraSAR-X and IKONOS-2) acquired over dense urban areas. Further, on basis of the proposed modifications, we also present a novel method to improve the sensor orientation of high resolution optical data (IKONOS-2) by obtaining ground control through local image matching, taking geometrically much more accurate TerraSAR-X images as a reference.

Apart from the joint application demands of SAR and optical imagery, the improved spatial resolution of SAR images from latest and future satellites like TerraSAR-X and TanDEM-X, is set to make a paramount impact on their usability. Here, the lack of any proven point feature detection and matching scheme for multisensor/multimodal SAR image matching encourages us to review the advancements in the field of computer vision and extend the applicability of Scale Invariant Feature Transform (SIFT) operator for SAR point feature matching. We have analysed the feature detection, identification and matching steps of the original SIFT processing chain. After thorough analysis, we propose steps to counter the speckle influence which deteriorates the SIFT operator performance for SAR images, in feature identification we evaluate different local gradient estimating techniques and highlight the fact that giving up the SIFT's rotation invariance characteristic increases the potential number of matches. In the feature matching stage we propose to combine MI and the SIFT operator capabilities for effective results in challenging SAR image matching scenarios. Further, our results indicate that a significant speedup is achieved on incorporating above suggested changes to the original SIFT processing chain.

ILLUSTRATIONS

List of Figures

List of Tables

LIST OF ABBREVIATIONS

BBF	Best Bin First
CRA	Cluster Reward Algorithm
DEM	Digital Elevation Model
GCP	Ground Control Point
GEC	Geocoded Ellipsoid Corrected
GIS	Geographic Information System
GPVE	Generalized Partial Volume Estimation
HS	High Resolution Spot Light
JH	Joint Histogram
MI	Mutual Information
NCC	Normalized Cross Correlation
NN	Nearest Neighbour
OE	Overlap Error
PCA	Principal Component Analysis
PVI	Partial Volume Interpolation
RC	Registration Consistency
ROEWA	Ratio of Exponential Weighted Averages
SAR	Synthetic Aperture Radar
SIFT	Scale Invariant Feature Transform
SPSA	Simultaneous Perturbation Stochastic Approximation
SURF	Speeded Up Robust Features
TAT	Turn around Time
TPE	Test Point Error
VHR	Very High Resolution

CHAPTER 1

INTRODUCTION

1.1 GENERAL

During the last decades, remote sensing sensors have undergone a rapid development in terms of both data quantity and characteristics. The Earth surface is now regularly being observed by sensors providing data in different wavelength domains, at different spatial and temporal resolutions. With this enormous increase in availability and quality of remote sensing data products, the use of remote sensing techniques for diverse applications has received a tremendous boost over the past few years. Remote sensing techniques have found applications in important areas like traffic studies, treaty and border monitoring, agricultural studies, generation of 3D models and topographic maps, early warning systems, urban growth, damage assessment and disaster mitigation. Specifically, remote sensing images have become one of the major sources of crisis information in the events of catastrophes and natural disaster. The practical utility of remote sensing data products and techniques can be seen from the activities of the centre for satellite based crisis information (ZKI) supported by the German Aerospace Centre (DLR)[1]. The chief objective of ZKI is rapid acquisition, processing and analysis of satellite data and the provision of satellite-based information products on natural and environmental disasters, for humanitarian relief activities, as well as in the context of civil security. In its short existence ZKI has utilized satellite imagery to generate invaluable information for decision makers in different catastrophic events like floods, hurricanes, oil spill, earthquakes, landslides and many more in different countries across the globe. For detailed information about the ZKI activities and the utility of remote sensing imagery in events of different natural disasters, the interested readers are encouraged to visit the ZKI homepage[2].

[1] www.dlr.de
[2] http://www.zki.dlr.de/intro_en.html

Normally, remote sensing applications need to accommodate images from different sensors/modalities; reason might be specific application demands or data unavailability. For example in case of a natural calamity, decision makers might be forced to use an old archived optical data with a newly acquired (post disaster) SAR image. Images acquired both by the passive optical sensors and active SAR sensors alone and in combination are major sources for crisis information management. In particular, the SAR sensors active nature gives them the capability to see through clouds and to acquire images at night which might be the only possible option during a catastrophic event. However, images acquired by SAR sensors have very different characteristics from normally used optical sensor images. On top of the very different geometry (sideways looking and measuring distances) from their optical counterparts (downward looking and measuring angles), images acquired by SAR sensors show a high amount of speckle influence caused by random backscatter of the microwaves emitted by the active sensor. Further, combined application of data from different sensors requires georeferenced and fine co-registered images for an accurate and successful analysis. Although, latest satellites provide the end user already georeferenced and orthorectified data products but still registration differences exist between various data sets acquired from different sources and agencies even after the orthorectification process. These differences need to be taken care of through quick automated registration techniques before using the images in different applications like cartographic mapping, change detection, image fusion, and 3D visualization and GIS overlays.

Image to image registration is a fundamental task in remote sensing image processing that is used to match two or more images taken, for example, at different times, from different sensors or from different view points. Typically, all remote sensing based applications require the registration of images as an intermediate step. Specific examples of applications where image registration is a significant component include (Brown, 1992):

i. Matching a target with a real time image of a scene for target recognition.
ii. Monitoring global land use and/or land cover through satellite images.
iii. Matching stereo images to recover shape for autonomous navigation.
iv. Integrating various remotely sensed images obtained from different sensors in different regions of the electromagnetic spectrum.

v. Inferring 3D information from images in which either the camera or the objects in the scene have moved.
vi. Multi-source classification.

Normally, the problem of registration problem can be divided into three overlapping categories:

i. **Multimodal registration:** Registration of images of the same scene acquired from different sensors. The problem here is to integrate images with different modalities, e.g. microwave and optical with an objective to benefit fusion and classification applications. Multimodal registration for remote sensing images also incorporates another category of registration known as the viewpoint registration as normally different modality remote sensing sensors also have a different view point or look angle during scene acquisition. Further, very different sensor acquisition principles and geometries (like very high resolution SAR and optical) can also complicate this type of registration.

ii. **Template registration:** To find a match of a reference pattern in an image. This kind of registration normally becomes useful for interpretation of well defined scenes such as airports, lakes etc, also useful to locate and identify well defined features and objects

iii. **Temporal registration:** Registration of images of same scene taken at different times or under different conditions. This category of registration has huge applications in planned change detection studies when same sensor images (mostly) are used for critical tasks of natural resource monitoring, surveillance of nuclear plants and urban growth monitoring.

1.2 PROBLEM STATEMENT

As stated earlier, remote sensing images are frequently used for a variety of applications such as image fusion (Pohl and Genderen, 1998), change detection (Dai and Khorram, 1998; Li et al., 2006), mosaic formations (Xin et al., 2007), traffic monitoring (Reinartz et al., 2006a), urban damage detection and mapping (Stramondo et al., 2007) and DEM generation (Hoja et al., 2006). Image to image registration is a prerequisite task in most of these applications, though the nature

and accuracy of registration may vary from one application to other. Currently, multimodal image registration is an established semi automatic process[3] in the form of a feature based technique available in commercial software like Erdas Imagine and ENVI. Human intervention is needed to select and match the tie points on the reference and the input images. Then, by taking into account the distortions that may be present in the images, a spatial transformation is estimated. The input image is then transformed according to the estimated transformation using a suitable resampling technique. Recently, Erdas Imagine from Leica Geosystems has introduced automated point feature matching capabilities but then it is still very much limited to same sensor images and performs matching using cross correlation and least squares method. Although, manual measurement of control points can sometimes achieve better accuracy to outperform an automated technique (Reinartz et al., 2009) but the pit falls of the described process lie in:

i. The availability of sufficient number of tie points, as some scenes may not have enough well defined features to use as tie points for an accurate registration.
ii. Time taken to identify and mark the tie points.
iii. The accuracy with which the points are selected and matched is dependent upon the operator's capability and experience.

Moreover, for latest remote sensing images from sensors like IKONOS and TerraSAR-X (Krieger et al., 2007), the elaborated process due to the limitations of human visual interpretation becomes much more labour intensive (especially in dense urban areas). The problem complexity can be visualized in Figure 1.1, depicting a road intersection imaged by IKONOS-2 and TerraSAR-X sensors. Due to fine minute details (like cars, small city structures) now available with the high resolution sensors, the idea of finding the same control/tie point (road intersections) becomes much more challenging than ever before. Although the fine minute details to a certain extent can be avoided by down sampling the images but then the control points should be marked accurately enough so that the resulting registration accuracy is in the order of originally available spatial resolution. Keeping in mind the shortcomings of the above process and the upcoming new challenges, the development of automatic image to image

[3] http://gi.leica-geosystems.com/documents/pdf/IMAGINEAutoSync_WhitePaper_Feb06.pdf

registration techniques especially suitable for high resolution imagery is encouraged. The need of automatic image registration is further strengthened by considering the fact that there has been a tremendous increase in data volume and in times to come processing large volumes of earth observation data is set to become a normal day activity.

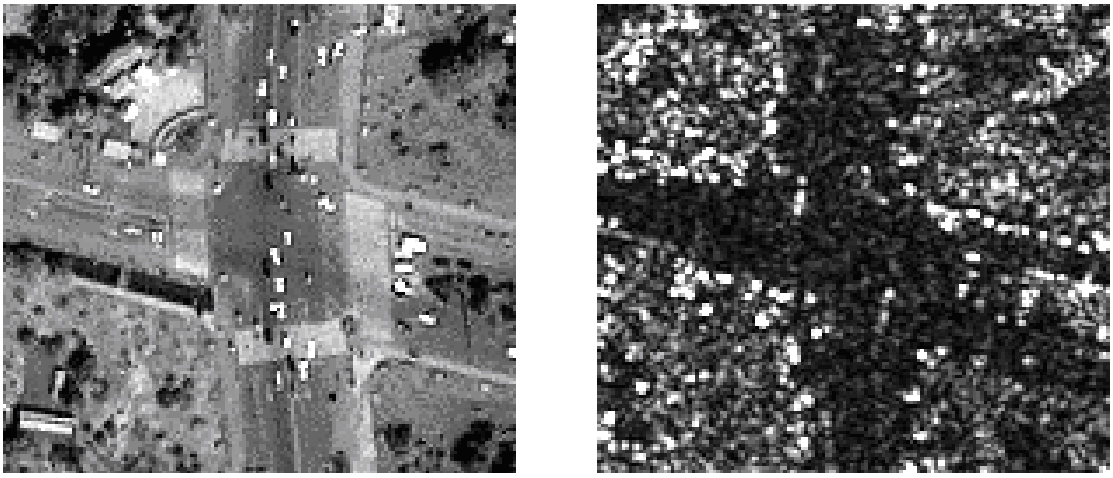


Figure 1.1: A road intersection in the City of Sichuan, China as imaged by (pan) IKONOS-2 (left) and (high resolution spot light) TerraSAR-X (right) sensors. High amount of minute details now visible with modern day sensors make the task of manually matching the control points more complicated than ever before (Suri and Reinartz, 2010). Images with 1 m pixel spacing are displayed at 175% zoom.

In general, a lot of work has been done in this field to develop registration schemes for optical images of similar spatial and spectral radiometric resolution. For registration of optical imagery both the control point detection and their subsequent matching have been automated by the remote sensing community over the past few years. Automated control point detection in optical imagery is very much possible using various edge (Canny, 1986) and corner detectors (Harris and Stephens, 1988; Förstner and Gülch, 1987). Further, for matching the already detected control points, techniques like least squares matching (Grün, 1985), cross correlation and moment invariants (Hu, 1962) have been successfully utilized. Lately, techniques from the field of computer vision in form of rotation and/or scale invariant feature detectors like SIFT (Lowe, 2004), Harris-Laplace (Mikolajczyk and Schmid, 2001), SURF (Bay et al., 2008) have proven their usefulness for detecting and matching homologous points in digital camera images (optical images acquired under lab conditions) but the applications of these techniques in the field of remote sensing image processing till date is very limited. In general, failure or inefficiencies of these mentioned techniques for images, specifically involving SAR-optical and

SAR-SAR image pairs necessitates further research and development in the field of remote sensing image registration.

Extracting common features from multimodal remote sensing SAR and optical images acquired with disparate acquisition principles, noise characteristics and sensor geometry, is a tedious task therefore intensity based registration techniques are preferred over the feature based ones. A lot of work has recently been reported on intensity based image registration, mostly concentrating on biomedical images (Pluim et al., 2003; Maes et al., 1997). Since, the characteristics of remote sensing images are different from biomedical images, the developed techniques may not be directly applicable and hence need to be suitably modified. For remote sensing images, similarity metrics namely cluster reward algorithm and mutual information have shown some effectiveness by showing capabilities of automatically aligning medium resolution images (5-15 m) through a rigid body transformation (Inglada and Giros, 2004; Chen et al., 2003a-b; Inglada, 2002). Still, their application to achieve registration amongst high resolution SAR and optical imagery having heterogeneous land cover classes (like urban, agricultural fields) needs to be thoroughly explored and analysed. The impact of distance measuring SAR sensor and the angle measuring optical sensors offer difficulties to intensity based metrics especially in urban areas and therefore certain modifications might be mandatory for their successful usage to high resolution imagery. Another important constraint related to the intensity based techniques is the registration turn around time for large remote sensing datasets which requires considerable attention before utilizing these metrics for any practical remote sensing application.

Reviewing feature based registration, till date there is no stable established technique for matching and registration of multimodal SAR image pairs. The inherent influence of multiplicative speckle makes feature detection and matching a complicated task. The complexities of different incidence and aspect angles during image acquisition offer further challenges to various established feature detectors and matching schemes. Looking into the advances in the field of computer vision, the scale invariant feature transform (SIFT) operator invented in (Lowe, 1999) and revised in (Lowe, 2004) by David Lowe is a feature detection algorithm which shows great potential for feature based registration of images with various

differences. The SIFT features are accredited to be invariant to image scale, rotation, illumination changes as well as are expected to offer high resistance to disruptions by occlusion, clutter or noise. For the mentioned reasons, SIFT operator has found immense applications in the fields of computer vision, pattern recognition and image processing (Battiato et al., 2007; Fiala and Shu, 2006). The inherent capability of the SIFT operator to match different resolution images (scale invariance) makes it a suitable alternative for diverse applications. Further, this property also offers a platform to select and match different characteristic features from a wide scale range. Therefore, for SAR images the SIFT operator can be utilized to detect and match stable features at lower scales, where the speckle influence is expected to diminish. Further, previous studies about comparative evaluation of the SIFT descriptors (for optical images) with other feature descriptors like shape context, PCA-SIFT, moments, differential invariants Steerable Filters have been proven to be very distinctive in the field of computer vision (Mikolajczyk and Schmid, 2005). The mentioned SIFT operator characteristics make it an attractive alternative to the intricate feature based multimodal SAR image registration problem.

1.2.1 Our Contribution

This dissertation caters to research in developing and analyzing methods for automated image registration techniques for multimodal (specifically high resolution SAR-optical and multimodal SAR) image pairs. Broad objectives of the study may be enumerated as:

i. Study and evaluation of intensity based registration metrics capable of registering SAR and optical images acquired by different sensors with different viewing geometries. To accomplish a detailed investigation of intensity based registration technique in terms of its applicability and efficiency for SAR and optical images.
ii. As the spatial resolution reaches to meter or sub-meter level it further widens the existing gap between SAR and optical sensor geometry and radiometry, posing problems for intensity based techniques. Modifications to the way the intensity based technique have been applied conventionally are proposed and analysed to extend their application for VHR imagery acquired especially over semi urban and dense urban land cover.

iii. Investigation of the scale invariant feature transform (SIFT) operator for solving the intricate multimodal SAR image registration problem. The operator is investigated and refined to enhance its performance for remote sensing SAR images. Detailed operator analysis in terms of its characteristics, applicability and efficiency for high and medium resolution SAR imagery is presented.

Specifically, on the basis of the evaluation of intensity based techniques for SAR-optical image matching, we present a novel approach to improve sensor orientation and orthorectification of high resolution optical data using the geometrical accuracy of TerraSAR-X data. A detailed analysis and evaluation of the results with available ground truth is presented. This exercise is targeted to highlight potential, issues and possible improvements pertaining to application of purely intensity based techniques, specifically for remote sensing applications.

1.3 ORGANIZATION OF THE DISSERTATION

The introduction has provided insights into the importance, applicability, issues and objectives of the research carried out for this dissertation. In the following chapter we extensively review the scientific literature to describe the state of the art in the field of remote sensing image registration. In Chapter 3, we elaborate intensity based image registration processing chain taking mutual information (MI) as the similarity metric. In the same chapter, we provide a thorough mathematic description of MI and also present its characteristics, relevant for its application in remote sensing. In Chapter 4, we propose modifications in conventional MI implementation to improve its applicability (accuracy and processing time) for high resolution multimodal imagery. In Chapter 5, we discuss two distinct approaches to register TerraSAR-X and IKONOS images acquired over dense urban areas. In the same chapter, results of a novel approach for orthorectification of high resolution optical data using TerraSAR-X imagery is also evaluated. In Chapter 6, we shift our focus to multisensor SAR image matching/registration, here we propose modifications in the originally proposed SIFT processing chain to improve its effectiveness for remote sensing SAR images. Finally, in accordance to the laid down objectives, we summarize our findings, discuss our accomplishments and advise guidelines for future investigations in this vital area of research.

CHAPTER 2

STATE OF THE ART

2.1 INTRODUCTION

The previous chapter laid the background of the problem envisaged in this dissertation. Automated multimodal image registration is an important problem as it is a pre-requisite in various diverse applications. As discussed, the solution to the problem has two approaches namely feature and intensity based techniques. The selection of one of the two techniques (alone or in combination) for image registration might finally depend upon the nature of the datasets involved. This chapter describes the registration problem formulation and reviews the already published work describing the state of the art in feature and intensity based techniques. Further, we explicitly highlight the recent advancements and current challenges being confronted in automating multimodal image registration.

2.2 IMAGE REGISTRATION

Image to image registration is a prerequisite in various applications in diverse fields and so has been an important area of investigation. Specifically, image registration has found immense applications in the fields of

i. Computer vision and pattern recognition
ii. Medical image analysis
iii. Remote sensing image processing

An extensive overview and survey of various image registration methods used in the above mentioned fields can be found in Brown (1992) and Zitová and Flusser (2003). Specifically, review for biomedical images (Pluim et al., 2003; Maintz and Viergever, 1998) and remote

sensing images (Wong and Clausi, 2007; Fonseca and Manjunath, 1996) can also be consulted. Mathematically, the problem of registering an input image I_I to a reference image I_R can be expressed as

$$I_R(x,y) = g\left(I_I\left(T(u,v)\right)\right) \tag{2.1}$$

where T is a transformation function which maps two spatial coordinates u and v, to the new spatial coordinates x' and y' (Equation 2.2) and g is a 1D intensity or radiometric interpolation function. Here it has to be noted that in Equation 2.1, transformation function T is normally estimated keeping only 2D features (no height information associated) into consideration as estimating transformation function using 3D features like topographic features, man made buildings etc. is a complicated task. Pictorially the problem has been depicted in Figure 2.1 but this representation is a very simple and abstract form of image registration which is seldom encountered in registering real world remote sensing images from contemporary sensors. A detailed discussion about this issue can be found later in this dissertation.

$$(x', y') = T(u,v) \tag{2.2}$$

The main objective of the entire process is to estimate the spatial transformation T and depending upon the method of resolution, image registration task can be divided into (Suri et al., 2009a):

Intensity based techniques: These register images largely on the basis of correlation between the intensity values of the pixels of two images (Inglada, 2002). This makes the registration process to emphasize on feature matching rather than on their detection. In this approach, the problem of registration is generally mapped as an optimization problem. Where the spatial transformation function T is the argument of the optimum of some similarity metric S, applied to reference image I_R and transformed input image I_{TI}. This can be expressed as

$$T = \arg(opt(S(I_R, I_{TI}))) \tag{2.3}$$

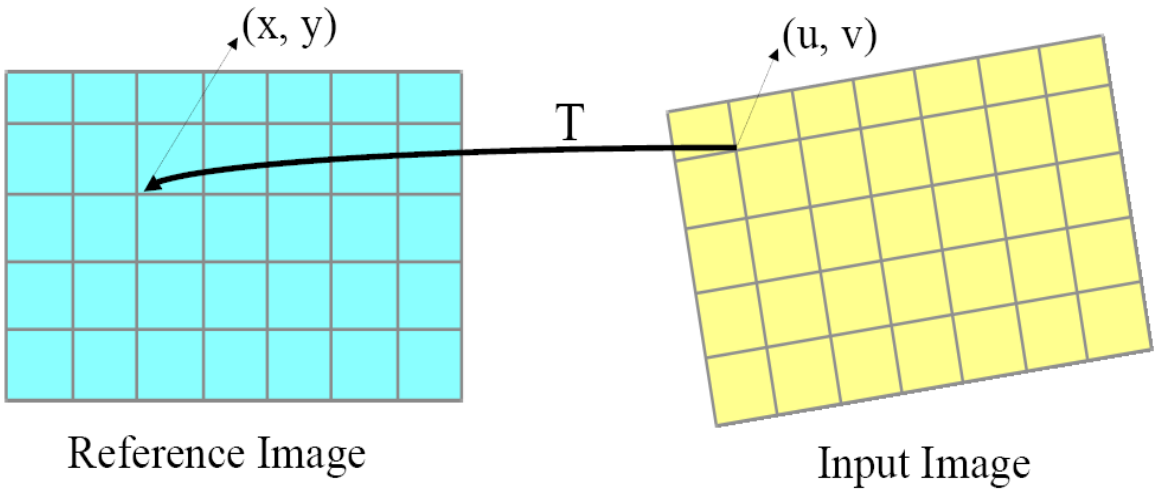


Figure 2.1: Pictorial depiction of the image registration problem. Solution lies in estimating the transformation function T that matches the (u, v) coordinate in input image to (x, y) coordinate in the reference image.

Considering the nature of implementation, these techniques might become sensitive to the changes in intensity values, introduced for instance by noise, by varying the illumination, and/or by using different sensor types. Intensity based registration techniques definitely have an edge over the feature based techniques in case of featureless areas where feature detection and matching might just not be possible. An intensity based registration processing chain has the following components:

i. Similarity metric
ii. Search space and strategy
iii. Transformation model and parameter estimation
iv. Image resampling and transformation
v. Registration quality assessment

Feature based techniques: These depend on the accurate identification of features or objects that describe important landmarks, sharp edges or shapes, which might be difficult to obtain. Significant regions (forests, lakes, fields), lines (region boundaries, coastlines, roads, rivers) or points (region corners, line intersections, points on curves with high curvature) may be considered as features. These represent information at higher level and are suitable for situations when changes in illumination are expected or if multi-sensor analysis is demanded. The

parameters of the spatial transformation T are computed from the extracted and matched feature in both the reference and the input images. The use of feature based techniques is recommended if images contain enough distinctive and easily detectable objects. The task of determining the best spatial transformation for the registration of the images can be broken down into the following major components:

i. Feature detection
ii. Feature matching
iii. Transformation model and parameter estimation
iv. Image resampling and transformation
v. Registration quality assessment

From the discussion until now it becomes clear that both feature based and intensity based techniques differ only in the first two steps of the registration process. In feature based registration techniques, explicit features need to be extracted and matched and on the contrary for intensity based techniques a suitable similarity metric along with a search strategy needs to be decided. The steps of transformation and model parameter estimation, image transformation and registration quality assessment are common to both the contrasting techniques. Hereafter, the advancements in all the registration steps in intensity and feature based registration techniques have been discussed.

2.3 INTENSITY BASED REGISTRATION

As already stated, intensity based registration techniques involve maximizing a suitable similarity metric between the images to achieve image registration (Equation 2.3). The current state of art in terms of utilized similarity metrics and search techniques is being provided here.

2.3.1 Similarity Metric

An important step to be considered in intensity based registration is the design or selection of a similarity metric to measure the similarity between the intensity values. The cost of evaluating

the similarity metric greatly influences the overall computational cost of the registration process. Some of the common similarity metrics that have been used for intensity based registration techniques are pixel averaging (Dasgupta and McGregor, 1992), sum of absolute errors (Turton et al., 1994), phase correlation (Reddy and Chaterji, 1996), correlation and normalized cross-correlation (Lewis, 1995), cluster reward algorithm (Inglada, 2002), mutual information (Chen et al., 2003a-b), normalized standard deviation (Woods et al., 1992) and correlation ratio (Roche et al., 2000). Information about each of the listed metrics can be obtained from the cited references. Out of these metrics, mutual information and cluster reward algorithm have been found capable of handling the multimodal SAR-optical image registration scenario. Mutual information in particular has also been extensively used for biomedical multi modal image registration (Pluim et al., 2003; Maes et al., 1997).

2.3.2 Search Space and Strategy

Having defined the similarity metric, image registration can be expressed as an optimization problem that aims to minimize a predefined cost function. Due to large computational costs associated with most registration problems, search space and strategy are key steps that require consideration. Search space is the class of transformations from which to find the optimal transformation to align the images. The type of transformation underlying each registration technique determines the characteristics of the search space and is based on the distortions and other variations present in the images. For example, if it is assumed that to register a pair of images, a translation must be performed, then the search space is the set of all translations over the range of reasonable distances. Normally, the search space of a registration problem is quite large and it is not possible to evaluate each transformation candidate. Therefore, it is important to reduce the number of computations. The greater the complexity of the misalignment between the images, the more severe this requirement will be.

A search strategy is employed to tackle the problem of a very large search space in registration problems. In large part, the choice of search strategy is determined by the characteristics of the search space including the form of the transformation (i.e. what types of constraints are to be satisfied? and how hard it is to find the optimum?). For example, if linear

inequalities are to be satisfied, then linear programming is advisable. If image features are composed of trees or graphs to be matched, then search strategies which are specialized for these data structures are required (Brown, 1992). Some of the common considerations to keep in mind before making a selection may be enumerated as:

i. Can the strategy take the benefits of parallel processing?
ii. Does the strategy make any assumptions?
iii. What is the storage and computational cost involved?

Examples of common search strategies include hierarchal or multiresolution approach, decision sequencing, relaxation, linear programming, tree and graph matching, dynamic programming and heuristic search. Description of these search strategies can be found in Brown (1992). Efficient search strategy forms the core component of an intensity based registration process. An efficient and effective search of registration parameters between two images has achieved some in recent past. Down sampling of images to form image pyramids to expedite mutual information based registration process has been demonstrated (Cole-Rhodes et al., 2003a; Chen and Varshney, 2000). The idea behind this setup is to get a rough estimate of the registration parameters from the coarser resolution images and iteratively improve the registration parameters by using finer resolution images in the formed image pyramids. The work presented by Cole-Rhodes et al. (2003a) enlists certain image pyramid formation technique and advocates the usage of Steerable Simoncelli Filters (Simoncelli and Freeman, 1995) more robust to translation, rotation and noise than the standard wavelet filters.

Intensity based registration techniques require a robust optimizer to achieve the elucidated task. Various optimizers with different mathematical backgrounds have been utilized to find accurate registration parameters between two images. Some of the optimization techniques that have been successfully employed in an intensity based registration process are Powell's method (Sun and Guo, 2007), simulated annealing (Zibaeifard and Rahmati, 2006), Nelder-Mead Simplex method (Xie et al., 2003), genetic algorithms (Bhattacharya and Das, 2007) and first/second order simultaneous perturbation stochastic approximation (SPSA) algorithm (Cole-

Rhodes et al., 2003a-b). For further reading, Pluim et al. (2003) also presented a survey of optimization techniques utilized in intensity based registration scenarios.

2.4 FEATURE BASED REGISTRATION

The advancements made in the critical tasks of feature detection and matching for any successful feature based registration (Equation 2.1) technique have been reviewed in this section.

2.4.1 Feature Detection

The feature detection is the most fundamental aspect of image registration for almost all image processing tasks. Feature detection extracts the information in the images that will be used for matching them. The information may be in the form of raw pixel values (i.e. the intensities). However, the other common feature spaces include (Brown, 1992):

i. Dimensionless features like corners of buildings and road intersections
ii. One dimensional features like edges, contours etc
iii. Two dimensional features like homogeneous closed boundary regions
iv. Three dimensional features like surfaces
v. Statistical features such as moment invariants or centroids

Conventionally, this step of image registration has been manual but some automation has been achieved in this critical phase. Features like line intersections (Stockman et al., 1982), high curvature points (Kanal et al., 1981), and centre of gravity of closed boundary regions (Goshtasby et al., 1986) have been successfully utilized but still reviewed literature suggests most of the remote sensing image registration applications have concentrated on using edges, contours and prominent points as features. Specifically for optical remote sensing images, feature detectors like the Förstner corner detector (Förstner and Gülch, 1987), Canny edge detector (Canny, 1986), Laplacian of Gaussian (Marr and Hilderth, 1980) and Harris corner detector (Harris and Stephens, 1988) have found tremendous applications for feature detection in image registration applications. As these operators are based on the assumption that the images suffer

from additive white noise these are not very successful for SAR images having multiplicative speckle influence (Bovik, 1988). For feature detection among SAR images (mainly edges) ratio based operators (Oller et al., 2002; Fjørtoft, et al., 1999, 1995; Touzi et al., 1988) have been developed and utilized for image registration applications.

2.4.2 Feature Matching

Once the features have been detected and identified, a mechanism for feature matching needs to be devised. Unambiguous representation of features known as feature descriptors have to be determined and finally matched using a functional relation also termed as a similarity metric. Techniques available to match two sets of features include template matching (Goshtasby, 1985), clustering (Stockman et al., 1982), relaxation (Ranade and Rosenfeld, 1980), matching of minimum spanning trees of the two sets (Lavine et al., 1983) and matching of convex hull edges of the two sets (Goshtasby and Stockman, 1985). Ideally, the feature descriptors and the matching scheme utilized should be scale and rotation invariant and should also accommodate for possible linear and non linear intensity variations in the feature space. For matching of detected edges in SAR and optical images, various edge matching techniques based on cross correlation (Oller et al., 2002), Hausdorff distance (Cheng et al., 2004) and contour models (Inglada and Adragna, 2001; Li et al., 1995) have been utilized with fair amount of success. Alternatively for point feature based techniques, remote sensing literature indicate successful usage of least square matching (Grün, 1985), moment invariants (Bentoutou et al., 2005) for feature matching. Further, some examples of extracting and matching prominent features like fields, lakes (Dare and Dowman, 2000) and buildings (Yang et al., 2005) can also be found.

2.5 COMMON REGISTRATION TASKS

In this section the registration tasks common both to intensity based and feature based techniques are being reviewed.

2.5.1 Transformation Model and Parameter Estimation

The type and parameters of the so-called mapping functions need to be decided for successful image to image registration. A transformation is defined as the mapping of location of points in one image to new locations in another image. For feature based registration techniques, transformation parameters are computed from the established feature correspondences. On the other hand, in the case of intensity based techniques the transformation function parameters are obtained through optimization of the similarity metric function computed mostly from the image intensity values. A transformation is normally selected depending upon the application needs and nature of distortions expected to be present within the images. Two types of transformation functions are commonly utilized in image registration:

Global transformations are implemented by a single equation which maps the entire image through one single function. Examples of global transformation are rigid, affine, projective, curved and polynomial transformations (Maintz and Viergever, 1998). Although, the global transformations are economical in terms of computational cost, these may not handle the local distortions. Distortions here refer to the noise and other deterministic effects that corrupt or alter the true intensity values and their locations in an image. Local distortions may arise due to differences in sensor viewpoint, change in subject position and other undesirable changes in the scene or sensor. Generally, the registration of satellite images has been modelled as a global deformation mainly utilizing affine and polynomial transformations (Bentoutou et al., 2005; Inglada and Giros, 2004; Chen et al., 2003). Global transformations which can accommodate slight local variations include surface splines (Rohr et al., 2001; Bookstein, 1989; Goshtasby, 1988a-b) and multiquadric transformations (Hardy, 1990). Among these transformation functions surface splines have the most local control while polynomial and weighted mean have the least local control. Multiquadrics allow some local deformation, but large local deformations result in large global deformations (Goshtasby, 1987). A sample application for thin plate spline (TPS) elastic wrapping for images in remote sensing images can be found in Xie et al. (2003).

Local transformations map the image differently depending on the spatial location and are thus much more difficult to express succinctly. To handle local distortions present within the images,

technique like piecewise mapping functions (Goshtasby, 1988b, 1987, 1986), surface spline mapping functions (Flusser, 1992; Goshtasby, 1988a) and elastic models (Brown, 1992) have proven their effectiveness. These transformations usually need large numbers of evenly distributed features to represent the local variation and are generally applied for high resolution images acquired from airborne sensors having complex local deformation caused by the wide view angle, terrain variation, the low flight height and the effects of yaw, pitch and roll (Devereux et al., 1990).

2.5.2 Image Resampling and Transformation

Since transformations are applied to the input image to resample and register it with the reference image, the use of an interpolation algorithm (Thévenaz et al., 2000) becomes an indispensable task. When a transformation is applied to the input image, a new grid is obtained and an intensity interpolation algorithm is necessary for the computation of new intensity values at every transformed grid point. Some of the most commonly used interpolation techniques are nearest neighbour, bilinear, cubic convolution (Keys, 1981), B-splines (Chaniotis and Poulikakos, 2004), cubic spline interpolation and radially symmetric kernels (Goshtasby et al., 2004). Several papers on interpolation techniques have been published in the past years. A detailed investigation and comparison of methods was carried out in Parker et al. (1983) for 2D images. Lehmann et al. (1999) also presented a survey of interpolation methods used in medical image registration applications.

2.5.3 Registration Accuracy Assessment

Estimation of registration accuracy achieved is a vital component of any registration process. Errors can be dragged into the registration process in each of the registration stages and also it is hard to distinguish between registration inaccuracies and actual physical differences in the image contents. Commonly utilized registration quality assessment tools in feature based registration techniques are localization error, matching error and alignment error (Zitová and Flusser, 2003).

Localization error: Displacement of the control points due to their inaccurate detection is called localization error. For practical cases this error cannot be measured or completely eliminated but can definitely be minimized by using optimal feature detection algorithms.

Matching error: Matching error is measured by the number of false matches in the feature matching step of the registration chain. This error can lead to the failure of a registration algorithm and should be strictly avoided by using robust matching algorithms. False matches can be avoided using consistency checks and cross validation techniques.

Alignment error: Alignment error denotes the difference between the mapping model used for the registration and the actual between-image geometric distortions. Reasons for alignment error include inaccurate mapping model utilized for registration or faulty computation of model parameters. The former case is caused by lack of a priori information about the geometric distortion while the latter originates from insufficient number of control points and/or their localization error. Alignment error can be measured by mean square error at the control points (CPE). Although, commonly used, it is not a good alignment error measure as it only quantifies how well the control point coordinates can be fitted by the chosen mapping model. Very similar to the CPE is so called the test point error (TPE). Test points are control points that are deliberately excluded from the calculation of the mapping parameters. TPE cannot become zero by over fitting making it more useful than CPE. This method can only be used if a sufficient number of control points are available. Otherwise, the exclusion of several control points may result in inaccurate estimation of mapping parameters.

Registration consistency: In absence of ground truth, registration consistency measure suggested by Holden et al. (2000) is useful for performance evaluation of intensity based registration techniques and is being discussed here. Let $T_{I,R}$ be the transformation obtained by maximizing the MI registration function by transforming the input image I over the reference image grid R. Similarly $T_{R,I}$ is the transformation obtained by maximizing the registration function by transforming the reference image R over the input image grid I. For two dimensional images, registration consistency RC is formulated as

$$RC = \left(\frac{1}{N}\right) \cdot \sum_{(x,y)\in(A_I \cap A_{I,R})} \left\| (x,y) - T_{R,I} \circ T_{I,R}(x,y) \right\| \tag{2.4}$$

where the composition $T_{I,R} \circ T_{R,I}$ is a result of applying $T_{R,I}$ followed by $T_{I,R}$. $A_{I,R}$ is the overlap area of image I and R respectively. A_I, A_R are their individual regions. N is the number of the pixels in the image that is being transformed. As the expression involves the composition $T_{I,R} \circ T_{R,I}$, cost associated with the described measure is inevitability of two way optimization of MI registration function. A two way optimization of MI may not be a big concern for small image sizes but for practical applications involving huge remote sensing data sets reporting of the achieved registration consistency may become a big overhead for the entire registration process. Important point to be kept in mind is that the value obtained from Equation 2.4 is not related to the ground level accuracy of registration achieved. The value obtained is a measure describing MI function consistency for image registration, in an ideal scenario value obtained from Equation 2.4 should be equal to zero. This measure was utilized for performance evaluation of intensity based registration of remote sensing images by Chen et al. (2003a-b).

Another approach to estimation of accuracy is consistency check using multiple cues. Here, the image registered by the method under investigation is compared (by a metric) with the same image registered by another comparative method. As the comparative method 'gold standard' method is believed to be the best in the particular application area or for the given image type. This approach is extensively used in medical imaging. In application areas where any gold standard does not exist, like in remote sensing, any other registration method can be taken as the comparative method. Finally, the oldest method of registration accuracy estimation, visual assessment by a domain expert should be mentioned. It is normally used as a complement of the mentioned objective error measures.

2.6 REGISTRATION IN THE FREQUENCY DOMAIN

Apart from achieving image to image registration by the above described methods registration schemes in the frequency domain have also been experimented and proposed in the past. In this section the registration techniques developed using properties of the Fourier transform are being

elaborated. By using the frequency domain, the Fourier methods achieve excellent robustness against correlated and frequency dependent noise. They are applicable, however only for images which have been at most rigidly misaligned. Registration techniques in the frequency domain depend upon a technique called phase correlation. Phase correlation is based on the Fourier Shift Theorem (Bracewell, 1965) and can be used for the registration of translated images. For a square image of size *NxN* pixels the discrete two dimensional transform is given by

$$F\left(\omega_1,\omega_2\right)=\frac{1}{N^2}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}f\left(x,y\right)^{e^{-j2\pi\left(\frac{x\omega_1}{N}+\frac{y\omega_2}{N}\right)}} \tag{2.5}$$

where $f\left(x,y\right)$ is the image in the spatial domain and the exponential term is the corresponding basis function of $F\left(\omega_1,\omega_2\right)$ in the Fourier space. According to the Fourier Shift Theorem, given two images f_1 and f_2 differing only by a displacement $\left(d_x,d_y\right)$ i.e.

$$f_2\left(x,y\right)=f_1(x-d_x,y-d_y) \tag{2.6}$$

their corresponding Fourier transforms F_1 and F_2 will be related by

$$F_2\left(\omega_x,\omega_y\right)=e^{-j(\omega_x d_x+\omega_y d_y)}F_1\left(\omega_x,\omega_y\right) \tag{2.7}$$

in other words, the two images have the same Fourier magnitude but a phase difference directly related to their displacement. The phase difference is given by $e^{J(\phi_1-\phi_2)}$. It turns out that if we compute the cross-power spectrum of the two images defined as

$$\frac{F_1\left(\omega_x,\omega_y\right)F_2^*\left(\omega_x,\omega_y\right)}{\left|F_1\left(\omega_x,\omega_y\right)F_2^*\left(\omega_x,\omega_y\right)\right|}=e^{\left(\omega_x d_x+\omega_y d_y\right)} \tag{2.8}$$

where F^* is the complex conjugate of F, the Fourier Shift Theorem guarantees that the phase of the cross-power spectrum is equivalent to the phase difference between the images. Furthermore, if we represent the phase of the cross-power spectrum in its spatial form, i.e. by taking the inverse Fourier transform of the representation in the frequency domain, and then we will have a function, that is approximately zero everywhere except at the displacement which is needed to optimally register the two images. Generally, Fourier transformation based registration is utilized to estimate global transformations within similar sensor images and especially is extremely useful for implementing cross correlation and convolution which are computationally expensive operations in the spatial domain.

In an extension of the phase correlation technique, DeCastro and Morandi (1987) proposed a technique to register images which are both translated and rotated with respect to each other. Further, numerous applications and improvements of the introduced technique for image registration are available in the literature (Stone et al., 2003; Foroosh et al., 2002; Reddy and Chatterji, 1996).The techniques in the frequency domain are fast and pretty stable for noisy images as the phase difference for every frequency contributes equally, the location of the peak will not change if there is noise which is limited to a narrow bandwidth. Consequently these techniques are useful for images acquired under different illumination conditions as illumination changes are very slow varying and therefore concentrated at low spatial frequencies. Along with the Fourier transformation, wavelet transformation has also been successfully utilized in image registration applications. Their primary applications have been feature extraction (Xie et al., 2008; Moigne et al., 2001; Li and Zhou, 1996; Djamdji et al., 1993) and generation of multiple representations of image data (Cole-Rhodes et al., 2003a; Stone et al., 1998; El-ghazawi et al., 1997; Moigne, 1995) to expedite the registration process.

2.7 RECENT ADVANCEMENTS AND CURRENT CHALLENGES

As already mentioned and highlighted, image registration has been an active area of research in the fields of remote sensing, computer vision and biomedical imaging. Thanks to advancements in research, a basic framework for both feature and intensity based techniques has been established but with the ever increasing remote sensing data volume and characteristics new challenges keep on arising and some of them have been answered in this thesis. Specifically the challenges and recent advancements in intensity and feature based techniques are as follows:

In intensity based registration, some similarity metrics like cross correlation, sum of absolute errors, mutual information, cluster reward algorithm and techniques based on least square matching have been found useful for image registration. In particular, mutual information and cluster reward algorithm have been found of some use to align remote sensing multimodal images automatically (Inglada and Giros, 2004). The major hindrance in using these similarity metrics for any practical purposes is the registration turn around time (Shu et al., 2005; Inglada, 2002). For large remote sensing datasets these metrics generally need huge amounts of time to

produce accurate registration results. A vigilant review of all the above mentioned work mainly done for 5-10 m spatial resolution imagery significantly indicates the usefulness of mutual information as a registration similarity metric. One sole example of combining an image classification step with mutual information for metric resolution imagery is found in Oller et al. (2006). Still the performance of mutual information for datasets from latest high resolution sensors like TerraSAR-X and IKONOS has not been explored to develop registration strategies for different industrial and academic applications. Moreover, the above mentioned published literature due to data unavailability during those times have not made any analysis of intensity based techniques for high resolution imagery acquired over urban areas where different sensor geometry and radiometry in combination complicates the registration task. Certain recent efforts in registering urban area images from airborne SAR sensors and optical satellite images like IKONOS have been made by combining feature and intensity based techniques (Lehureau et al., 2008; Wegner et al., 2008; Yang et al., 2005). But considering the complicated nature of SAR imagery, registration techniques dependent upon line and point detectors are not a robust alternative as till date here has not been any proven feature detector, capable of extracting conjugate features in SAR and optical images.

In the area of feature based registration, there is still no established technique that can guarantee conjugate features amongst images with various differences. Feature based SAR image registration is a difficult task due to the presence of multiplicative speckle influence (Touzi et al., 1988; Bovik, 1988). Still, some work in the field of feature based SAR image registration can be found in the remote sensing literature. Borghys et al. (2001) presented an approach utilizing digital topographic maps for registration of high resolution polarimetric SAR images. A point feature based approach was used successfully for registration of SAR-SAR and optical-optical images by Bentoutou et al. (2005). Here, the authors utilized the Harris operator (Harris and Stephens, 1988) for detection of control points (strong corner points). Subsequently, a template matching algorithm based on invariants based similarity measure was used. Specifically, for multimodal SAR images techniques based on edge maps have been experimented in Li et al. (1995). Edge or contour based registration techniques are generally time consuming involving heavy computation (Inglada and Giros, 2004) and depend heavily on prominent edges to be successfully detected and matched which might become complicated due

to the presence of speckle or absence of features having strong edge characteristics. Ideally, for this scenario robust feature detectors not biased to any of the standard geometrical shape needs to be explored. Further, the feature identification through descriptors and their robust matching is also an area of concern. Lately, techniques from the field of computer vision in form of rotation and/or scale invariant feature detectors like SIFT (Lowe, 2004), SURF (Bay et al., 2008), Harris-Laplace (Mikolajczyk and Schmid, 2001) have proven their usefulness for detecting and matching homologous points in digital camera images (optical images acquired under lab conditions) but their application in the field of remote sensing image processing, till date is very limited. Therefore, the application of the SIFT descriptors for SAR image registration and matching applications like Radargrammetry (Tupin and Nicolas, 2002) has been suitably analysed in this presented work.

2.8 SUMMARY

This chapter completes an exhaustive review of the extensive literature available in the field of image registration. Image registration has applications in different fields and so has been extensively researched by researchers with various perspectives in different domains. The advancements in feature based, intensity based and frequency based registration methodologies have been accordingly cited and acknowledged. The chapter also formulates the current state of art in context of remote sensing multimodal image registration according to the literature reviewed. In the following chapter, we proceed with the basics of intensity based registration followed by their application to various registration scenarios.

CHAPTER 3

INTENSITY BASED REGISTRATION

3.1 INTRODUCTION

The task of automating multimodal image registration has been a topic of research for decades now. Specifically in the field of remote sensing, availability of new sensors generating images with ever improving spatial, spectral and radiometric details keeps on throwing fresh challenges to existing image registration methodologies thus instigating new research and development in the field. Considering the meticulous task of extracting and matching features in very different SAR and optical images, the development of a universally applicable feature based technique is difficult. Exploring the area or intensity based techniques the concepts like normalized cross correlation, least squares matching and sum of absolute differences are only effective for images with similar modalities. Moving from these conventional techniques, in this chapter, we present information theory based similarity metric namely mutual information, which has capabilities to handle multimodal images. Henceforth, we present in detail the mathematical properties, implementation details and characteristic properties of mutual information in context of registering discrete multimodal remote sensing images.

3.2 NORMALIZED CROSS CORRELATION

We start our analysis by first highlighting the limitation of the extensively used cross correlation for registration of images acquired by SAR and optical sensors. Cross correlation is a method utilized in signal processing to determine the correlation between two signals. This concept with a basic statistical approach has been extended in image processing and can be utilized for feature matching and registration applications. The idea in both the cases is to compute the correlation between two images by a template matching technique. This method is generally useful for images, which are misaligned by small shifts in both horizontal and/or vertical directions. Cross

correlation as a measure of similarity between an image I and an image template T positioned at (u,v) is defined as (Lewis, 1995)

$$cc(u,v) = \sum_{x,y} I(x,y)T(x-u,y-v) \tag{3.1}$$

Further, a normalized cross correlation (NCC) having values in range [-1, +1] and defined in Equation 3.2 is mostly utilized in various applications:

$$ncc(u,v) = \frac{\sum_{x,y}\left[I_{x,y} - \overline{I}_{u,v}\right]\left[T(x-u,y-v) - \overline{T}\right]}{\sqrt{\sum_{x,y}\left[I(x,y) - \overline{I}_{u,v}\right]^2 \sum_{x,y}\left[T(x-u,y-v) - \overline{T}\right]^2}} \tag{3.2}$$

Where $\overline{T}$ and $\overline{I}_{u,v}$ are the means of T and $I(x,y)$ respectively. This metric computes pixel-wise cross correlation and normalizes it by the square root of the image auto correlation function and has been considered a much more robust measure of alignment as compared to the cross correlation measure defined in Equation 3.1. Normalized cross correlation as represented in Equation 3.2 in not invariant to scale, rotation and perspective distortions. We present the performance of the normalized cross correlation for registration of a sample SAR and optical imagery.

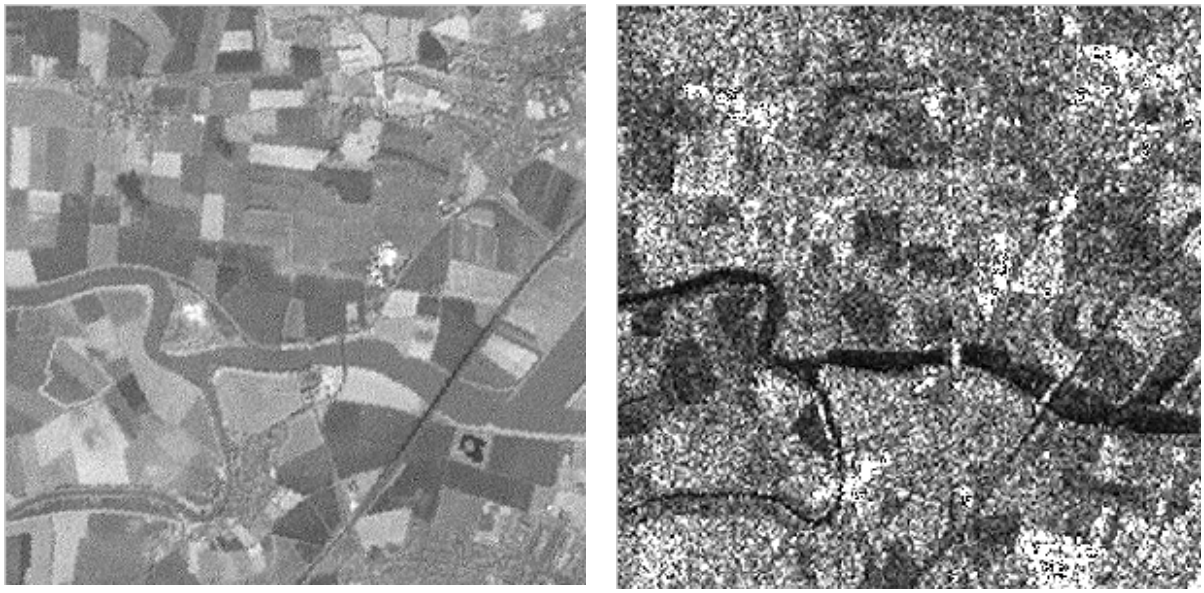
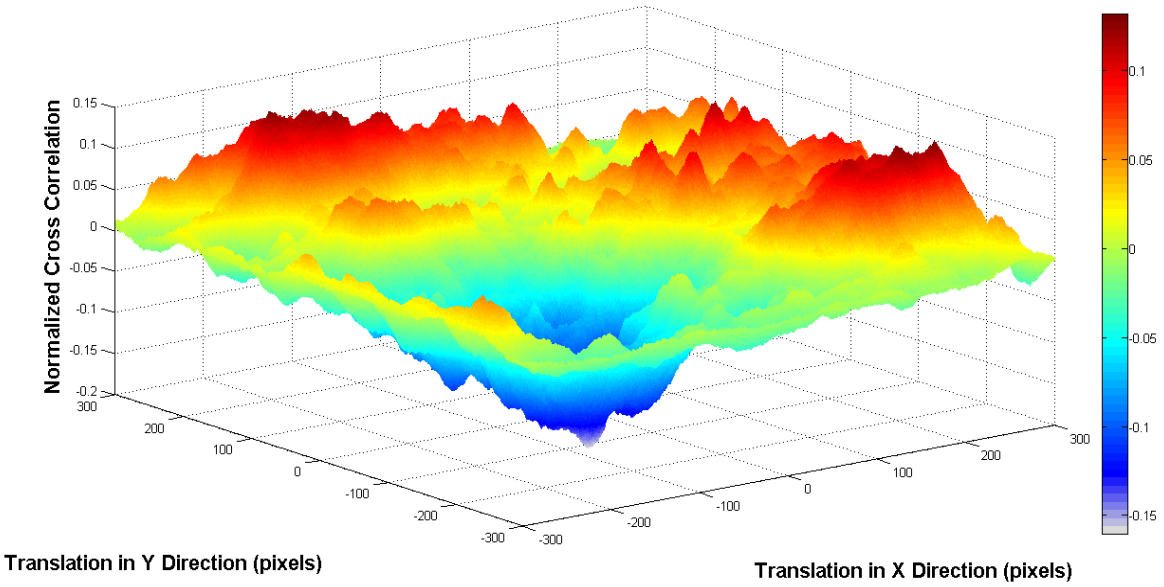
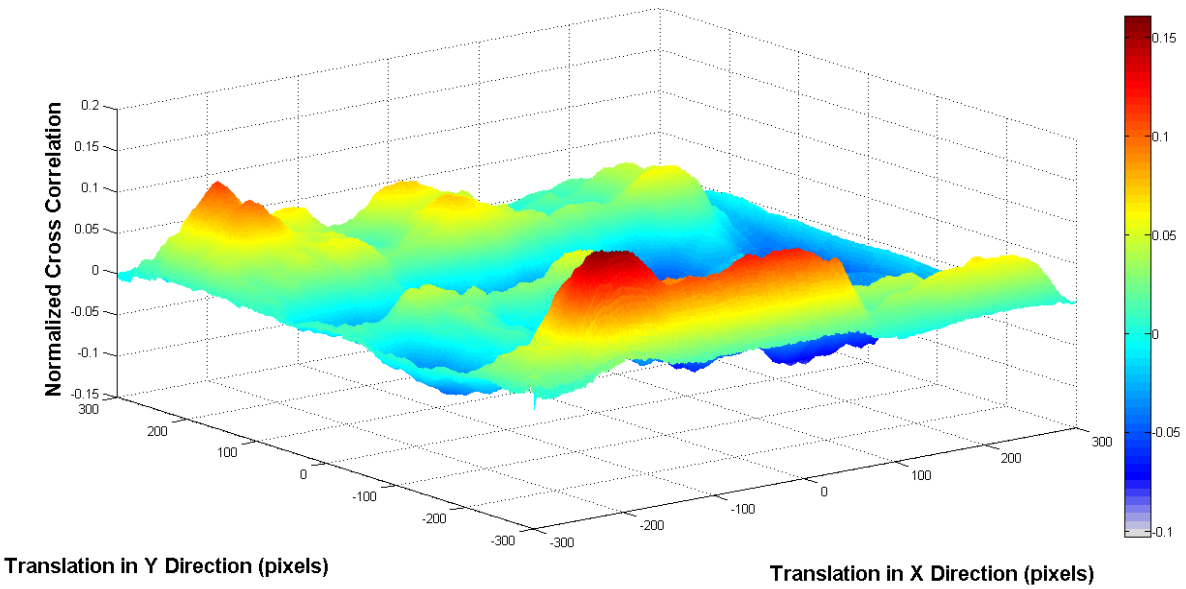


Figure 3.1: Pre-registered Landsat (left) and Radarsat (right) images utilized to test similarity metrics.



(a)



(b)

Figure 3.2: Performance of NCC for multimodal image registration. (a) Landsat image used as template and Radarsat image used as the reference. (b) Radarsat used as template and Landsat image used as the reference. Failure of NCC for multimodal image pairs asks for investigating other similarity metrics for automatic image registration.

Figure 3.1 depicts Radarsat and Landsat images which have already been registered manually for analysis purposes. Both the images have a pixel spacing of 14.25 m and the image size is 301 x

301 pixels. As the images are pre-registered a similarity metric (NCC) peak at position (0, 0) representing no misalignment between the images is expected. The performance of NCC for the images depicted above can be visualized in Figure 3.2, where we interchangeably utilize Landsat and Radarsat images as template and reference respectively. It is observed that in both the cases the NCC failed to produce a peak at perfect alignment of the two pre-registered images.

Considering its mathematical construction, the failure of NCC for SAR optical matching is very much justifiable. NCC by concept needs a linear relationship between input and the reference image and therefore it is observed that correlation coefficient based similarity measures are limited to handling images acquired with the same modality. The relationship between the multimodal SAR and optical intensity values is definitely much more complex and therefore asks for a much advanced similarity metric discussed in the following sections[4].

3.3 INFORMATION THEORETIC MEASURES

Information theory explores the fundamental limits of the representation and transmission of information. Further, it also attempts to characterize the information present in random variables providing a vital background for researchers in the areas of data compression, signal and image processing, controls, and pattern recognition. It was developed out of Shannon's pioneering work in the 1940's at Bell Laboratories (Shannon, 1948). His work focused on characterizing information for communication systems by finding ways of measuring data based on the uncertainty or randomness present in the given system. Shannon proved that for probabilities p_i,

$$\sum_i p_i \log p_i \tag{3.3}$$

is the only functional form that satisfies all the conditions that a measure of uncertainty should satisfy. Entropy is one of the main building blocks of information theory. From it are obtained two other major building blocks, relative entropy and mutual information. Entropy is a measure of uncertainty (or information) in a random variable, relative entropy is a distance measure between one probability distribution and another, and mutual information is the amount of

[4] In general two similarity metrics capable of handling SAR optical registration are cluster reward algorithm (CRA) and mutual information (MI). However, in this dissertation we focus primarily on MI as it has been found much more robust in comparison to CRA (Suri and Reinartz, 2009).

information that one random variable contains about another (Cover and Thomas, 1991). As already mentioned information theory has applications in various fields and hence its basic building blocks can be explained with different perspectives according to the underlying application. As the sought application for this dissertation comes under digital image processing therefore all the discussions and concepts of information theory and related building blocks have been presented in the context of discrete digital images. For a more general and a complete understanding of the wide applications of information theory interested readers are referred to a text book on elements of information theory (Cover and Thomas, 2006).

3.3.1 Entropy and Information

In digital imaging, an intensity image X is represented as a matrix of intensity values. For an n bits/pixel image, the intensity values are discrete greyscale values, $\aleph = \{x_1, x_2, ...x_N\}$ where $N = 2^n - 1$. A histogram can be constructed from an image by looking at each pixel intensity value and counting the number of times a pixel intensity value occurs, or the number of times a pixel intensity value lies in a range. Dividing the histogram of occurrences by the total number of pixels in the image gives the frequency of occurrence of each intensity value, or each intensity value bin. Normalizing the histogram in this way gives an estimate of the probability distribution function of pixel intensity values for the image. Given an image X, we use p to denote the corresponding estimated intensity value probability distribution function, where $p(x) = \mathrm{P}\left(\mathrm{X_{i,j}} = x\right)$, for $x \in \aleph$ and $X_{i,j}$ is a pixel in image X. A result of histogram normalization is the following

$$\sum_x p(x) = 1 \tag{3.4}$$

The function p is normally referred as probability density function or the probability mass function. Entropy uses probability distribution or density function to measure the randomness or uncertainty of a random variable. Under the assumption that each observation in the image matrix is independent and occurs with the probability determined by the frequency of occurrence, the entropy of the random variable X, or the entropy of the image, can be computed.

The entropy $H(X)$, for a discrete random variable X, with probability distribution function p, is defined as

$$H(X) = H(p) = -\sum_{x \varepsilon \aleph} p(x) \log p(x) \tag{3.5}$$

where for reasons of continuity the assumption $0 \log 0 = 0$ is made. The notation $H(p)$ emphasizes the dependence of entropy on the probability distribution of X, as opposed to the actual intensity values of X.

3.3.2 Joint Entropy and Mutual Information

For defining the concept of joint entropy and mutual information, two discrete images A and B over their regions of overlap can be considered as observations from two discrete random variables, A and B, with probability distributions p and q. It is also assumed that the two random variables A and B have sample spaces α and β respectively. The 2D joint histogram can be constructed from images A and B over their region of overlap by counting the number of times the intensity pair (a,b) occurs in corresponding pixel pairs $\left(A_{i,j}, B_{i,j}\right)$. Normalizing the joint histogram gives an estimate of the joint probability distribution r, where $r(a,b) = P_r\left(A_{i,j} = a, B_{i,j} = b\right)$. A result of the joint histogram normalization is that $\sum_a \sum_b r(a,b) = 1$. The marginal of the joint distribution are defined as

$$\sum_{a \varepsilon \alpha} r(a,b) = q(b)$$

$$\sum_{b \varepsilon \beta} r(a,b) = p(a) \tag{3.6}$$

Joint entropy $H(A,B)$ is a functional of the joint probability distribution r, and is a measure of the combined randomness of the discrete random variables A and B. Further, the joint entropy can be defined as

$$H(A,B) = H(r) = -\sum_{a \varepsilon \alpha} \sum_{b \varepsilon \beta} r(a,b) \log r(a,b) \tag{3.7}$$

If two random variables are independent, then the joint probability distribution becomes the product distribution d, i.e. $r(a,b) = d(a,b) = p(a)q(b)$. In this situation, joint entropy simplifies to

$$H(A,B) = -\sum_{a,b} r(a,b)\log r(a,b)$$

$$= -\sum_{a,b} p(a)q(b)\log p(a) - \sum_{a,b} p(a)q(b)\log q(b)$$

$$= H(A) + H(B) \tag{3.8}$$

In general $H(A,B) \le H(A) + H(B)$, with equality if and only if A and B are independent.

Mutual information in terms of conditional entropy: Conditional entropy measures the randomness of a variable B given A, and is given by

$$H(B \mid A) = \sum_{a\in\alpha} p(a)H(B \mid A = \alpha)$$

$$= -\sum_{a\varepsilon\alpha} p(a)\sum_{b\varepsilon\beta} p(b \mid a)\log p(b \mid a)$$

$$= -\sum_{a\varepsilon\alpha}\sum_{b\varepsilon\beta} p(b,a)\log p(b \mid a)$$

$$= H(B,A) - H(A) \tag{3.9}$$

MI can be expressed using Equation 3.9, the relation in that case quantifies dependence within two variables

$$MI(A,B) = H(B) - H(B \mid A) \tag{3.10}$$

using Equation 3.9

$$MI(A,B) = H(B) + H(A) - H(B,A) \tag{3.11}$$

Mutual information in terms of relative entropy: MI can also be defined as a special case of relative entropy. It measures the amount of information shared between two random variables, or the decrease in randomness of one random variable due to the knowledge of another. Relative entropy, or Kullback-Leibler distance, is a measure of the distance between one probability distribution and another. It measures the error of using an estimated distribution q over the true

distribution p (Cover and Thomas, 1991). The relative entropy $D(p \| q)$, of two probability distributions p and q over $\aleph$, is defined as

$$D(p \| q) = \sum_{x \in \chi} p(x) \log\left(\frac{p(x)}{q(x)}\right) \tag{3.12}$$

where for reasons of continuity, we define $0 \log(0/q) = 0$ and $p \log(p/0) = \infty$ .With A and B be two random variables with probability distributions p and q, respectively, and joint probability distribution r. Mutual information, $MI(A,B)$, is the relative entropy between the joint probability distribution r, and the product distribution d, where $d(x,y) = p(x)q(y)$. That is

$$\begin{aligned} MI(A,B) &= D(r \| d) \\ &= \sum_{a \varepsilon \alpha} \sum_{b \varepsilon \beta} r(a,b) \log\left(\frac{r(a,b)}{p(a)q(b)}\right) \end{aligned} \tag{3.13}$$

As mentioned earlier, if the random variables A and B are independent, then the joint probability distribution is equal to the product distribution, i.e. $r = d$. Thus, mutual information measures the correlation between A and B, with respect to A and B being independent.
The relation expressed in Equation 3.11 is depicted by a Venn diagram in Figure 3.3.

$$MI(A,B) = H(A) + H(B) - H(A,B) \tag{3.14}$$

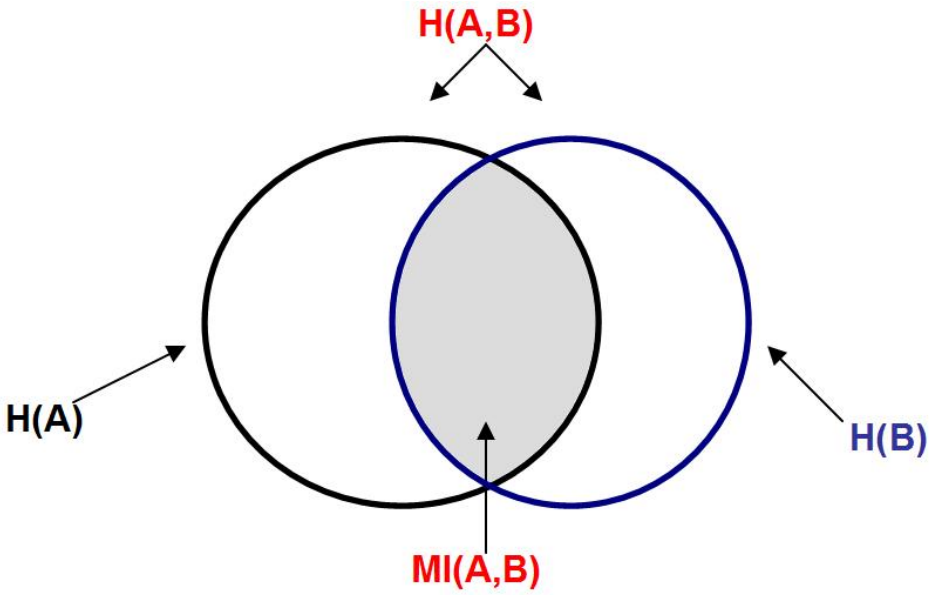


Figure 3.3: Relationship between variables marginal entropies, joint entropy and mutual information.

The two image entropies are represented as the two sets. The union of the two entropies represent the joint entropy of the system and the intersection of the two sets is equivalent to the mutual information contained within the two images. In this dissertation, image registration is achieved by maximizing this mutual information content within images.

3.3.3 Statistical Properties of Mutual Information

i. $MI(A,B) = MI(B,A)$

MI formulation in Equations 3.9 and 3.12 are derived using two different definitions. From both the definition sources the symmetry of mutual information can be inferred. In practice, for registration applications, MI function might lose its symmetric nature due to the artefacts from different processing chain components.

ii. $MI(A,A) = H(A)$

The information image A contains itself is equal to information of image A

iii. $MI(A,B) \le H(A); MI(A,B) \le H(B)$

The information the images or random variables contain about each other can never be greater than the information in the images itself.

iv. $MI(A,B) = 0$, if and only if A and B are independent

Normally, MI function attains strictly positive values. If it attains a value of zero for two random variables it signifies that knowledge of A does not give any information about B. Implying, A and B are not related in any way.

3.3.4 Mutual Information Definitions

We utilize the concepts from information theory (mutual information) to solve the problem of multimodal image registration especially involving SAR and optical images. Mutual information as a similarity metric for multimodal image registration has gained a lot of importance in recent

past. It was first and independently used as a measure for matching of medical images by Viola (1995) and Collignon et al. (1995). Since its introduction, MI without any mandatory pre-processing has been rigorously utilized for image registration applications. Mutual information depending upon the underlying entropy measure can be defined in following ways:

Rényi entropy: The Rényi entropy, a generalization of Shannon entropy, is one of a family of functions for quantifying the diversity, uncertainty or randomness of a system and is defined as follows:

$$R_\alpha(A) = \frac{1}{1-\alpha} \ln \sum_{i=1}^{n} p^\alpha(x_i) \qquad \alpha \in \Re - 1 \tag{3.15}$$

The Rényi entropy is additive in nature, it becomes the Shannon entropy as $\alpha \to 1$ and hence for this case mutual information is defined as

$$MI_\alpha^R(A,B) = R_\alpha(A) + R_\alpha(B) - R_\alpha(A,B) \tag{3.16}$$

Tsallis entropy: Tsallis entropy is a one-parameter generalization of Shannon entropy, has been often discussed in statistical physics as a new information measure. It is generally defined as:

$$S_q = \frac{1}{q-1} * \left(1 - \sum_{i=1}^{k} (p_i)^q\right) \tag{3.17}$$

where k is the total number of entities and the real number q is an entropic index that characterizes the degree of non-extensivity. As Tsallis entropy is non additive in nature, mutual information in this case is defined as

$$MI_q^s(A,B) = S_q(A) + S_q(B) + (1-q) S_q(A) S_q(B) \tag{3.18}$$

Mutual information based upon different entropy calculation has been tested for multimodal image registration scenarios (Wachowiak et al., 2003) but the most extensively used measure is the one based on Shannon entropy and is also utilized here for all further discussions.

Shannon entropy: As discussed earlier, by nature Shannon's entropy measure is additive in nature, implying the fact that if random variables A and B are independent, then mutual information between two variables is defined as

$$MI(A,B) = H(A) + H(B) - H(A,B) \tag{3.19}$$

where $H(A)$ and $H(B)$ are the Shannon entropies of A and B respectively, $H(A,B)$ is the joint entropy and $H(A|B)$ is the conditional entropy of A given B. Shannon entropy of a variable x is defined as

$$H(x) = \int_{-\infty}^{+\infty} f(x) \ln f(x) dx \tag{3.20}$$

where $f(x)$ is the probability density function of x. Registration of two images A and B is based on maximization of $MI(A,B)$ (Equation 3.19), which means that the joint entropy $H(A,B)$ is minimized. At the same time, the sums of the marginal entropies $H(A)$ and $H(B)$ are maximized to avoid situations where the images are shifted so far apart that only two small regions of background coincide. Marginal entropies ensure that the overlapping region of images contains most of their information. During discrete implementations because of the MI formulation in Equation 3.19 the metric can produce negative values specifically when images are shifted far apart, to tackle this issue and to reduce the sensitivity of MI towards changes in image overlap, the following definition of mutual information (Studholme et al., 1999) is utilized

$$MI(A,B) = \frac{H(A) + H(B)}{H(A,B)} \tag{3.21}$$

For implementation purposes, the entropies and joint entropy can be computed from (Chen et al., 2003a-b)

$$H(A) = \sum_{a} -p_A(a) \log p_A(a) \tag{3.22}$$

$$H(B) = \sum_{b} -p_B(b) \log p_B(b) \tag{3.23}$$

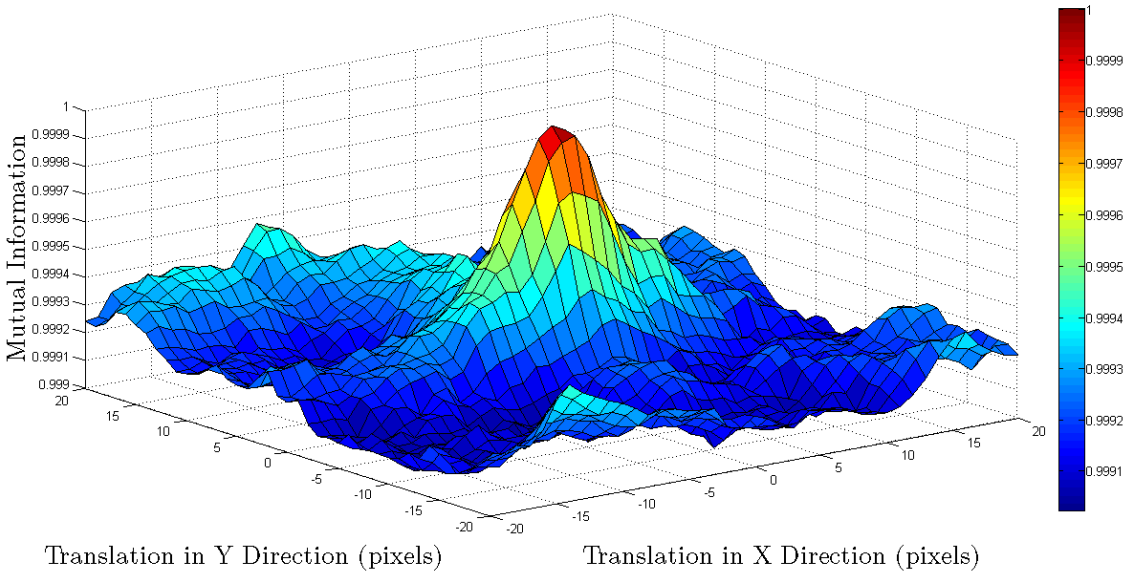
$$H(A,B) = \sum_{a,b} -p_{A,B}(a,b) \log p_{A,B}(a,b) \tag{3.24}$$

$$p_{A,B}(a,b) = \frac{h(a,b)}{\sum_{a,b} h(a,b)} \tag{3.25}$$

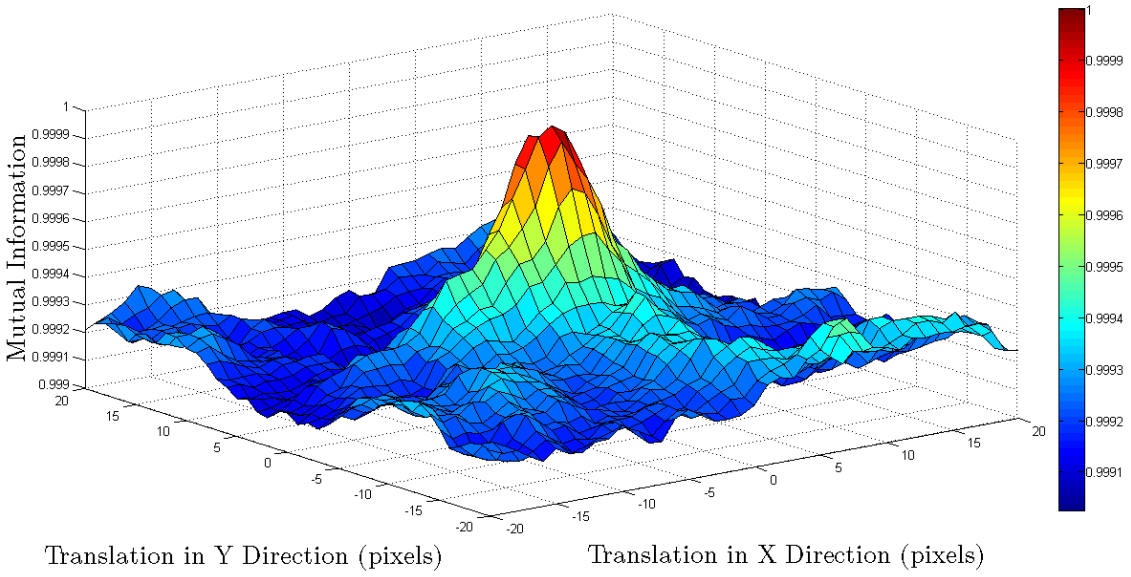
$$p_A(a) = \sum_{b} p_{A,B}(a,b) \tag{3.26}$$

where h is a joint histogram (JH) of the two images involved. It is a 2D matrix with the intensity values of one image along one axis and the intensity values of the other image along the other

axis. The value $h(a,b)$ is the statistic number of corresponding pairs having intensity value a in the first image and intensity value b in the second image.



(a)



(b)

Figure 3.4: Performance of MI for multimodal image registration. (a) Landsat image used as input and Radarsat image used as the reference. (b) Radarsat image used as input and Landsat image used as the reference. In both the cases MI detects the desired peak (0, 0) for the pre-registered images.

Thus, it can be seen from Equations (3.22) to (3.26) that the joint histogram is the only requirement for MI computation between any two images. Therefore, different joint histogramming techniques and influence of joint histogram bin size on MI performance is thoroughly discussed later in this chapter. A sample mutual information performance for the Radarsat and Landsat imagery (Figure 3.1) both interchangeably used as reference and input images can be seen in Figure 3.4. For both the graphs the input image has been translated over the reference image grid from [-20 +20] pixels in both x and y direction. The registration peak for the two images is observed at perfect alignment of the two images.

3.4 REGISTRATION STEPS

Generally, a typical intensity based registration scheme requires the following steps:

i. *Spatial Transformation:* The type and parameters of the mapping functions need to be decided for successful image to image registration.
ii. *Radiometric Normalization:* Mathematical explanation of MI highlights that it is heavily dependent upon the joint histogram of the images being registered. Therefore, the joint histogram (JH) size needs a vigilant selection.
iii. *Joint Histogramming Technique:* To estimate the joint histogram of the transformed input image and the fixed reference image, suitable interpolation techniques need to be incorporated in the process.
iv. *Search Space and Strategy:* The problem of image to image registration finally reduces to searching the optimal set of transformation from a large search space that is capable of aligning the images perfectly.

Each of the above briefly mentioned steps would now be discussed in details.

3.4.1 Spatial Transformations

A transformation is defined as the mapping of location of points in one image to new locations in another image. A transformation is normally selected depending upon the application needs and

nature of distortions expected to be present within the images. As already mentioned in Chapter 2, on the basis of transformation nature, two types of transformation functions are commonly utilized in the field of image registration:

i. Global transformation functions
ii. Local transformation functions

In particular global transformations are generally utilized for satellite image registration, and high resolution aircraft images having complex local deformation caused by the wide view angle, terrain variation, the low flight height and the effects of yaw, pitch and roll are registered using local transformation (Desheng et al., 2006). The presented work concentrates on remote sensing satellite images with much stable geometries and thus the utilized global transformation models are being elaborated. It is to be noted here that single global transformation might still not be sufficient for an entire complete remote sensing scene but here for analysis we consider smaller areas extracted from full coverages.

Rigid transformation

For remote sensing registration applications, the most commonly used are the rigid body/affine and polynomial transformation functions. Global affine incorporates geometric operations of scaling, translation and rotation. It typically has four parameters tx, ty, s, θ which map co-ordinates (X_I, Y_I) of input image to (X_R, Y_R) of the reference image as follows

$$\begin{pmatrix} X_R \\ Y_R \end{pmatrix} = \begin{pmatrix} t_x \\ t_y \end{pmatrix} + s \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} X_I \\ Y_I \end{pmatrix} \tag{3.27}$$

This can be rewritten as

$$\overline{p_R} = \bar{t} + sR\overline{p_I} \tag{3.28}$$

Where $\overline{p_R}$, $\overline{p_I}$ are the coordinate vectors of the reference and the input images, $\bar{t}$ is the translation vector, s is the scale factor and R is the rotation matrix. Since the rotation matrix R is orthogonal (the rows or columns are perpendicular to each other) the angles and lengths in the original image are preserved after the registration. Because of the scale factor s, the rigid body

transformation allows change in length relative to the original image, but it is the same in both x and y directions.

Affine transformation

The general 2D affine transformation is represented as

$$\begin{pmatrix} X_R \\ Y_R \end{pmatrix} = \begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} X_I \\ Y_I \end{pmatrix} \tag{3.29}$$

The above representation does not have the properties associated with the rigid body transformation described by Equation 3.28. Angles and lengths in this case need not be preserved but parallel lines do remain parallel. This transformation can also account for more general spatial deformations like shear and changes in aspect ratio. Affine transformation represented in Equation 3.30 are global transformations and are sufficient to match two images of a scene taken from the same viewing angle but from a different position, i.e. the camera can be moved and it can be rotated around its optical axis (Brown, 1992).

Polynomial transformation

Polynomial transformations are one of the most general transformation models normally utilized for image registration. Achieving image registration using polynomial transformations assumes some combination of the following constraints (Goshtasby, 1988):

i. The scene is flat.
ii. The area to be imaged is small compared to the distance of the camera to the scene.
iii. Viewing angle is the same when the images are obtained.
iv. The camera is a perfect pin-hole camera.
v. The effect of atmospheric turbulence or other distortions in the images is negligible.

A polynomial transformation of degree M is defined by

$$X_R = \sum_{j=0}^{M} \sum_{k=0}^{j} a_{jk} X_I^{\,k} Y_I^{\,j-k} \tag{3.30}$$

$$Y_R = \sum_{j=0}^{M} \sum_{k=0}^{j} b_{jk} X_I^{\,k} Y_I^{\,j-k} \tag{3.31}$$

Parameters a, b are determined by feature correspondences in the reference and the input image. The minimum number of feature correspondences for first and second order polynomial transformations are three and six respectively. Generally, more number of features (N) then required for parameter estimation are utilized and parameter of the mapping function are determined by minimizing the sum of squared errors (Equation 3.32)

$$E = \sum_{n=1}^{N} \left[\left(X_{R_n} - \sum_{j=0}^{M} \sum_{k=0}^{j} a_{jk} X_{I_n}^{\,k} Y_{I_n}^{\,j-k} \right) + \left(Y_{R_n} - \sum_{j=0}^{M} \sum_{k=0}^{j} b_{jk} X_{I_n}^{k} Y_{I_n}^{\,j-k} \right) \right]^2 \tag{3.32}$$

Generally, guidelines for polynomial transformations advocate well spread out control points, sufficient in number for a second or third order polynomial estimation. Usually, polynomials of degree larger than 2 are not advised for image geometric transformation because prediction of their behaviours from their coefficients gets difficult (Goshtasby, 1986). Abundant examples for usage of global polynomial transformations for image registration are available in the literature. Some of them are included in Lehureau et al. (2008), Panigrahi and Tripathy (2002), Woods et al. (1998).

As already mentioned, the selection of a spatial transformation between the two images depends solely on the nature of deformations present. For example, in case of remote sensing images orthorectified (with or without using a DEM) to same coordinate projection system generally differ mainly by two translations in x and y direction respectively. Here it is assumed that the scale factor between the two images being registered can be accurately estimated from the image ground spacing distance. Therefore, a scenario where we have to co-register two orthorectified scenes, selection of two translation parameters for the optimization process should suffice the purpose. In case, images are in different coordinate systems, a much more complex transformation like affine or polynomial transformation might be inevitable.

3.4.2 Joint Histogramming Techniques

For computing the MI registration function the joint histogram of the transformed input image and the reference image needs to be estimated. As depicted in Figure 3.5, according to the selected transformation function, the transformed input image may or may not coincide with the target reference image grid. Therefore, in certain cases of grid misalignment an exact joint histogram may not be obtained and certain approximation becomes inevitable. For joint histogram estimation, one step and two step histogramming techniques have been utilized in the past (Chen and Varshney, 2004b).

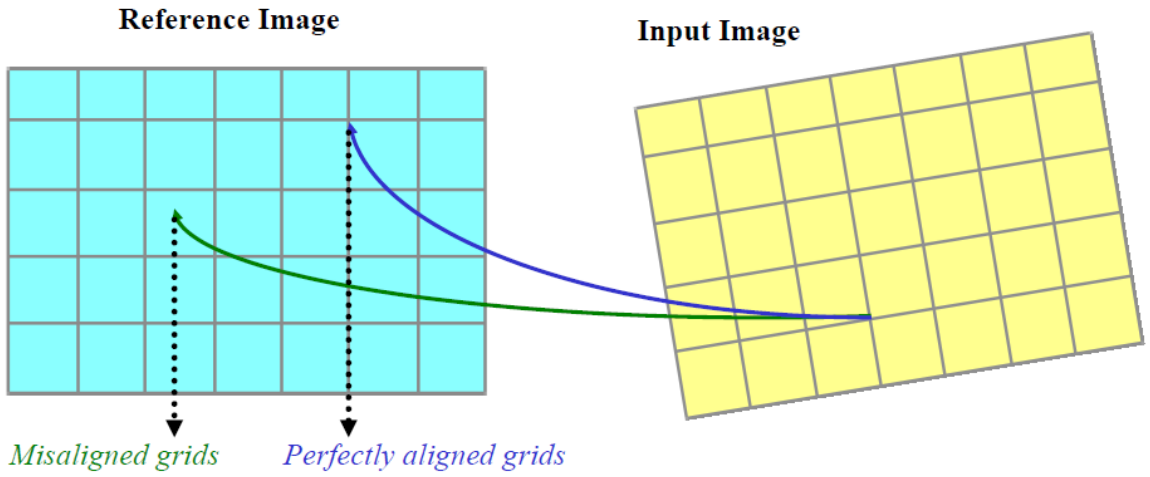


Figure 3.5: Image transformation in image registration. Depending upon the spatial transformation function selected the transformed input image may or may not coincide with the reference image grid.

Two step techniques evaluate the intensity values of the reference image at the transformed grid points through interpolation and than update a unique joint histogram entry by one. Interpolation techniques like nearest neighbour, bilinear interpolation, cubic convolution and cubic B-spline interpolations have been utilized for mutual information based registration methods (Chen and Varshney, 2002). In general, this intensity value interpolation for updating a single joint histogram entry has been found to introduce a systematic regular pattern in MI registration function which can make the registration function noisy, in turn making the registration optimization process a tedious task. These regular patterns in the literature of MI based registration are generally termed as interpolation induced artefacts (Pluim et al., 2000).

Figure 3.6 depicts induced artefacts for nearest neighbour, bilinear and cubic convolution intensity interpolation techniques. MI function in Figure 3.6a-c have been plotted by sliding the Landsat image over the Radarsat image (Figure 3.1) in a range of [-10 +10] pixels in x direction with a 0.5 pixel movement. A regular pattern especially in the case of bilinear and cubic convolution interpolation (Figure 3.6 b-c) is observed where the MI registration function attains local minima at exact grid alignments (no interpolation) and local maxima at sub pixel image movements (interpolation step). Due to this mentioned recursive behaviour the MI function attains a peak at 0.5 pixel misalignment for the perfectly registered images.

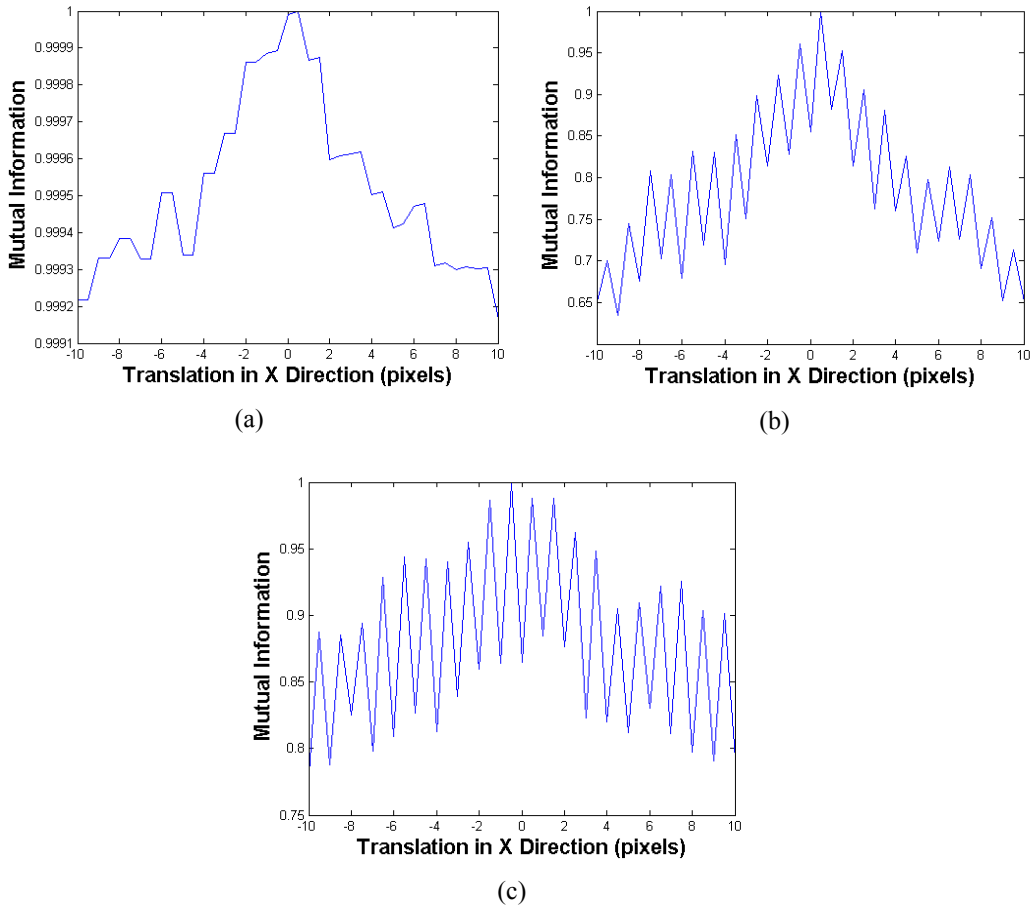


Figure 3.6: Interpolation induced artefacts appear in mutual information registration function with two step joint histogramming techniques. (a) Nearest neighbour, (b) Bilinear interpolation, (c) Cubic convolution.

In case of nearest neighbour interpolation the artefacts do not make a very regular pattern but this interpolation technique can sometimes make the MI function immune to sub pixel

movements (blurring effect). This can be observed in Figure 3.6a where MI function value does not change for a consecutive pixel and sub pixel image translation at several locations in a small search space of 41 x-axis translations. This observed behaviour is not a desirable metric characteristic as sub-pixel registration accuracy is desired in lot of remote sensing applications. The problem of interpolation induced artefacts has been addressed by lot of researchers in different fields (Inglada et al., 2007; Tsao, 2003; Pluim et al., 2000). Two reasons have been mainly highlighted to produce interpolation induced artefacts:

Pluim et al. (2000) mention equal sample spacing in one or more dimensions as one of the prime reasons. For example, if two images have the same sample spacing along the horizontal axis, then any integer values horizontal displacement between the two images results in the alignments of all horizontal grid lines within the overlap area. Therefore, fewer interpolations are required to estimate the joint histogram, than in the case where grids are not aligned (sub pixel movements). This phenomenon is observed for sample in Figure 3.6b as both the images have same pixel spacing in the horizontal direction and MI function shows a regular pattern for pixel and sub pixel movements.

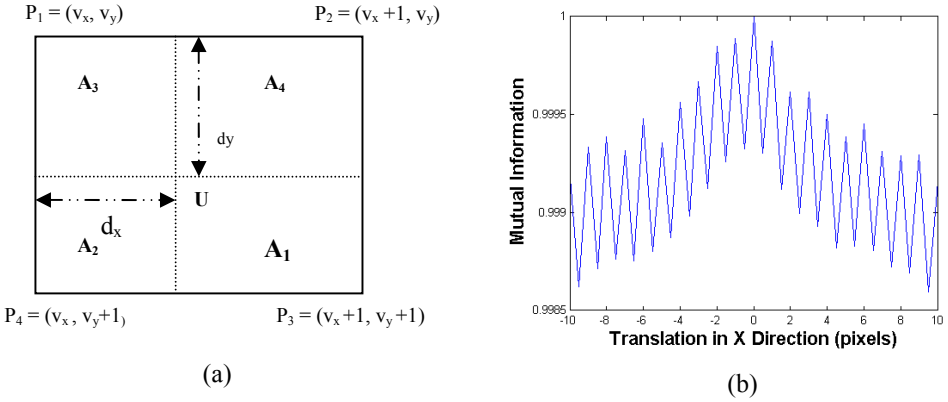
Chen and Varshney (2004) also indicate noise within the images as one of the factors generating artefacts in mutual information surface. In the cited reference, the authors with a registration experiment involving Landsat TM (band 1 and band 3) images showed that addition of different levels of Gaussian noise in both the images produced artefacts in MI search spaces. Moreover, it was also highlighted that severity of the artefacts increased with the variance of the added Gaussian noise. Observations from this optical image registration experiment can be safely extended to SAR-optical registration scenario as different statistics and levels of noises in both the images can very well influence the smoothness of the registration function.

Considering the persistent problem of interpolation induced artefacts one step joint histogramming techniques have been proposed and utilized. One step techniques update the joint histogram without resorting to the interpolation of intensity values at the transformed grid points on the reference image. These techniques include the partial volume interpolation (PVI) suggested by Maes et al. (1997) and later extended by Chen and Varshney (2003) to generalized

partial volume estimation (GPVE) which utilize B-spline interpolation to update several joint histogram bins as opposed to a single joint histogram entry done by the two step techniques.

Partial volume interpolation

As compared to the standard interpolation techniques which generate an interpolated transformed input image to update a single joint histogram entry, partial volume interpolation does the following to handle the two cases of grid alignment. For the case of perfect grid alignment nothing is performed and a single joint histogram entry is updated in the joint histogram. In case of grid misalignment, PVI updates four entries in the joint histogram by the same weights as for linear interpolation. Figure 3.7a presents a graphical representation of the scenario of an input grid point U transformed by a transformation function to an unaligned grid point U' in the reference image grid.



(a) (b)

Figure 3.7: (a) Graphical illustration of Partial Volume Interpolation (PVI). (b) Behaviour of mutual information registration function using the PVI scheme. (For images shown in Figure 3.1).

The PVI updates four joint histogram bins comprising the nearest four reference image grid points surrounding the transformed grid point U. The four bins in the joint histogram are updated according to their respective weights calculated according to the rules of linear interpolation. The entire step is performed as

$$h\left(R\left(P_i\right),I\left(U\right)\right)=h\left(R\left(P_i\right),I\left(U\right)\right)+A_i \qquad \text{i} = 1,2,3,4 \tag{3.33}$$

Here, h represents the joint histogram of the reference image R and the transformed input image I. A_i represent the respective areas of the four reference image grid points surrounding the transformed input image grid point U'. A sample performance of the discussed partial volume interpolation scheme for images shown in Figure 3.1 (Landsat Radarsat image pair) can be visualized in Figure 3.7b. This scheme introduced in the field of medical imaging did not recover completely from the problems of interpolation induced artefacts. We again observe the regular pattern and the important point to be noted here is the fact that here the regular pattern is just opposite of the pattern obtained using the two step bilinear interpolation (Figure 3.6b). in this case we obtain a MI peak at perfect alignment of the two images but it is also not a desirable behaviour as in this case the registration function starts taking local maxima at perfect grid alignments and local minima at misaligned grid positions.

Generalized partial volume estimation (GPVE)

Partial volume interpolation updates one histogram entry for perfectly aligned grids and four histogram entries for misaligned grid transformations, thus producing a regular dispersion in the joint histogram of the images being registered. This regular dispersion when combined with the above explained artefact generating condition lead to the undesirable PVI behaviour obtained in Figure 3.7b. Citing these reasons Chen and Varshney (2003) extended the concept of partial volume interpolation to generalized partial volume estimation (GPVE). GPVE is also a one step joint histogramming technique and has been utilized for various results and discussions in this thesis and is elaborated thoroughly in this section.

Let I and R be the input and the reference image respectively and T be a transformation that is applied to input image grid points. Considering the Figure 3.7a and assuming that T maps the grid point $\left(u_{x'}, u_{y'}\right)$ in image I onto the point with coordinate, $\left(v_x + d_x, v_y + d_y\right)$ in the reference image R, where $\left(v_x, v_y\right)$ is a grid point in R and $0 \le d_x, d_y < 1$. GPVE is mathematically described as

$$h(R(v_x + p, v_y + q), I(u_{x'}, u_{y'})) += f_1(p - d_x) * f_2(q - d_y)^{5} \tag{3.34}$$

In the equation above f_1 and f_2 refers to the kernel function selected, p and q specify the pixels involved in the joint histogram updating procedure depending upon the support of the selected kernel functions f_1 and f_2. The selected kernel function f must satisfy the following conditions:

i. where x is a real number

ii. $\sum_{n=-\infty}^{\infty} f(n-d) = 1$, where n is an integer $0 \le d < 1$ (3.35)

Normally, B-spline functions fits in the role of the kernel functions mentioned in Equation 3.36. Basis splines (B-splines) are one of the most commonly used family of spline functions. These are derived by several self convolutions of a basis rectangular function.

$$h_1(x) = \begin{cases} 1, & 0 \le |x| < 0.5 \\ 0, & \text{elsewhere} \end{cases} \tag{3.36}$$

A B-spline of degree n or order $n+1$ can be derived through n self convolutions of the function described in Equation 3.36.

$$h_N(x) = \underbrace{h_1(x) * h_1(x) * \ldots h_1(x)}_{\text{N - 1 times}} \tag{3.37}$$

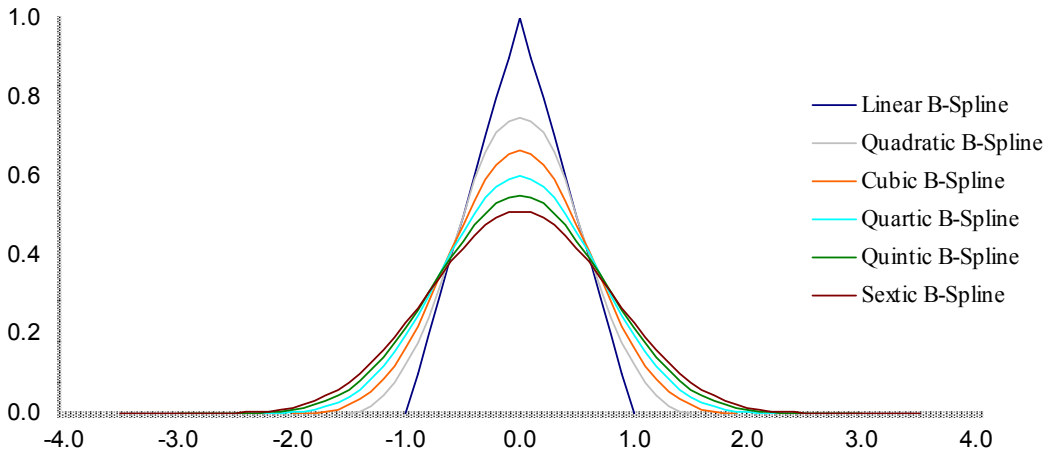


Figure 3.8: Different B-spline kernels utilized in GPVE.

[5] A+=B implies A=A+B

On selecting B-splines as kernel functions for GPVE, the support of the underlying B-spline kernel (depicted in Figure 3.8) decides the number of grid points that are involved in the joint histogram updating procedure for the two mentioned scenarios (exactly aligned grid and misaligned grids). For B-spline kernels (order 2 to 7 depicted in Figure 3.8), Table 3.1 tabulates the number of grid points involved in joint histogram updating procedure for the two cases. A pictorial description of grid points involved in updating the joint histogram in case of aligned and misaligned grids for B-spline kernels of order 2-4 is provided in Figure 3.9. The number of histogram bins being updated for every pixel transformation is $O\left(n^2\right)$ where n is the order of underlying B-spline kernel. In general, all the B-spline kernels in their discrete form can be implemented with the help of look up tables.

Table 3.1: B-spline kernel behaviour as interpolating functions in GPVE (Suri and Reinartz, 2008).

Order	Aligned Grids	Misaligned Grids[6]	Ratio
Linear (2)	1	4, 2	4.00
Quadratic (3)	9	6, 4	2.25
Cubic (4)	9	12, 16	1.78
Quartic (5)	25	20, 16	1.56
Quintic (6)	25	30, 36	1.44
Sextic (7)	49	42,36	1.36

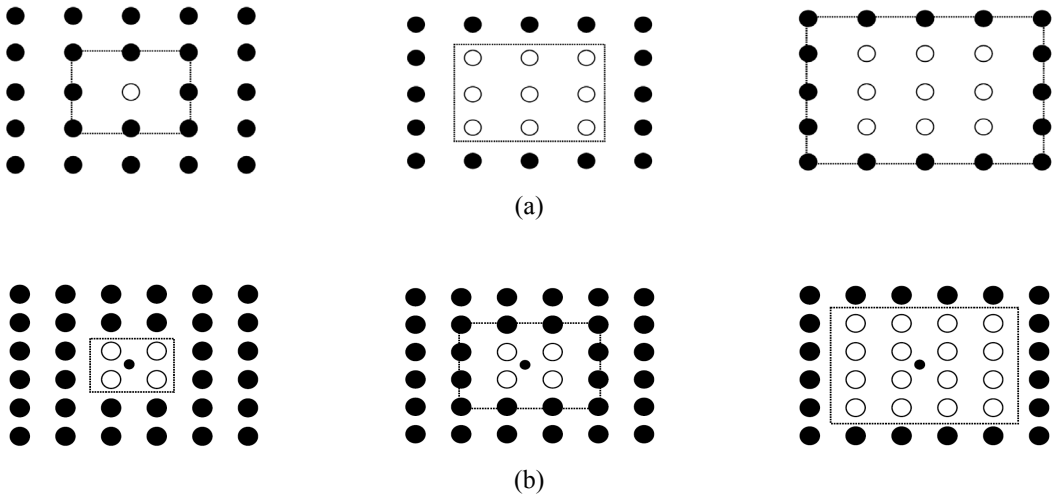
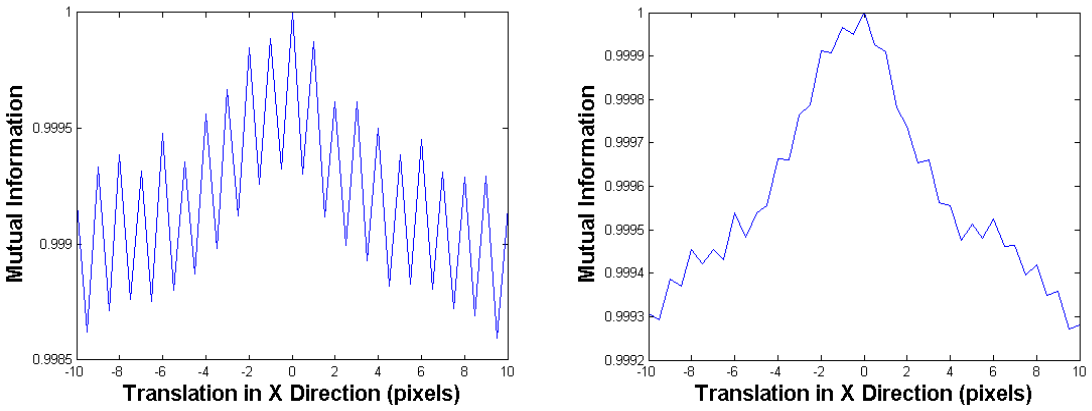


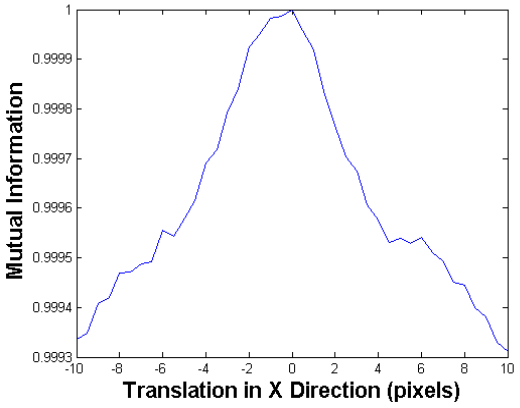
Figure 3.9: Grid points that are involved in updating the joint histogram in case of GPVE. (a) Transformed grid point coincident with the reference grid point. (b) Transformed grid point is surrounded by the reference grid points. Figures are shown only for B-spline kernels of order 2, 3 and 4. The rectangular block depends upon the support of the underlying B-spline kernel (Figure 3.8).

[6] Depending upon the nature of misalignment

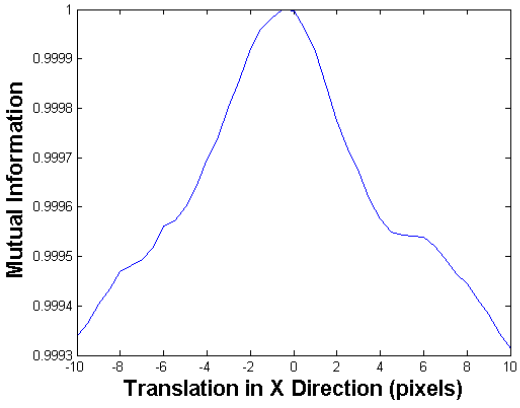


B-spline order 2

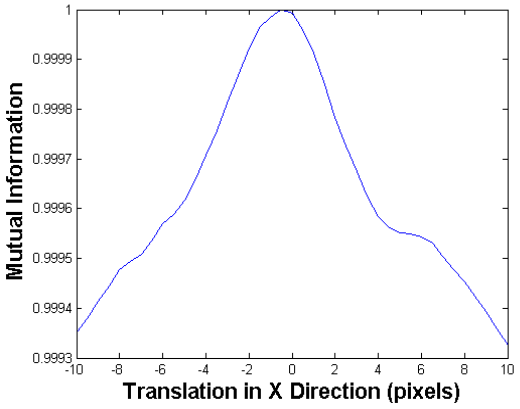
B-spline order 3



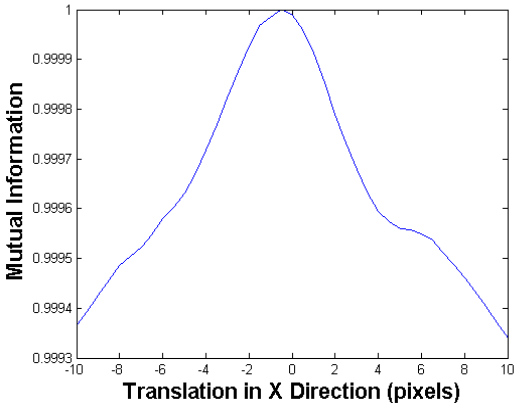
B-spline order 4



B-spline order 5



B-spline order 6



B-spline order 7

Figure 3.10: Performance of the GPVE technique using B-spline kernels from order 2 to 7.

The ratio of maximum to minimum number of updated entries in the joint histogram updating procedure (max-min ratio), a factor reducing the artefacts appearance in the MI surface is

tabulated in Table 3.1. As expected, the max-min ratio tends to 1 as order of interpolating B-spline kernel increases leading to a lesser joint histogram dispersion which is considered as a reason to produce interpolation induced artefacts in mutual information registration function. Further, this fact can also be analysed from the MI graphs obtained for the Landsat-Radarsat image pair depicted in Figure 3.1. We show the MI behaviour with respect to the underlying B-spline function in Figure 3.10. It can be clearly observed that order 2 B-spline kernel behaves very similarly to the PVI technique producing systematic interpolation artefacts (both these function have a max-min ratio of 4). Further, as we increase the support of the B-spline kernel from order 2 to order 7 to reduce the joint histogram dispersion through the max-min ratio (listed in Table 3.1) the interpolation artefacts disappear and the MI function becomes smoother and free of any regular artefact patterns.

3.4.3 Radiometric Normalization

Considering the radiometric resolution of modern day sensors like TerraSAR-X (16 bit) and IKONOS (11 bit), radiometric normalization becomes mandatory for any application of the joint histogram based mutual information. For example, a scene from TerraSAR-X may have intensity values in the entire possible dynamic range of [0, 65535] (16 bit radiometry) and the corresponding scene from IKONOS-2 may have intensity values in range [0, 2047] (11 bit radiometry). Here, the JH size without any intensity binning would be an array of size 65536 x 2048, leading to memory and computational issues. In practice, scenes from sensors having high radiometric resolution may not show a good distribution of intensity values over the whole histogram and therefore would most of the times result into sparse joint histograms, adding inefficiency to the registration process. This fact can be very well be visualized in Figure 3.11 where pre- registered images of size 451x451 pixels from TerraSAR-X and IKONOS sensors have been depicted with their respective histograms. It is observed that most of the information content in both the images is located in a very small fraction of their respective intensity range. In this case, most of the IKONOS image information is contained in the range [68, 426] (Figure 3.11c) and most of the TerraSAR-X image information is contained in the range [0, 342] (Figure 3.11d). This uneven distribution of the information content in both the images leads to a very spare joint histogram of size 1030x4263, depicted in Figure 3.11e. To handle the problem of

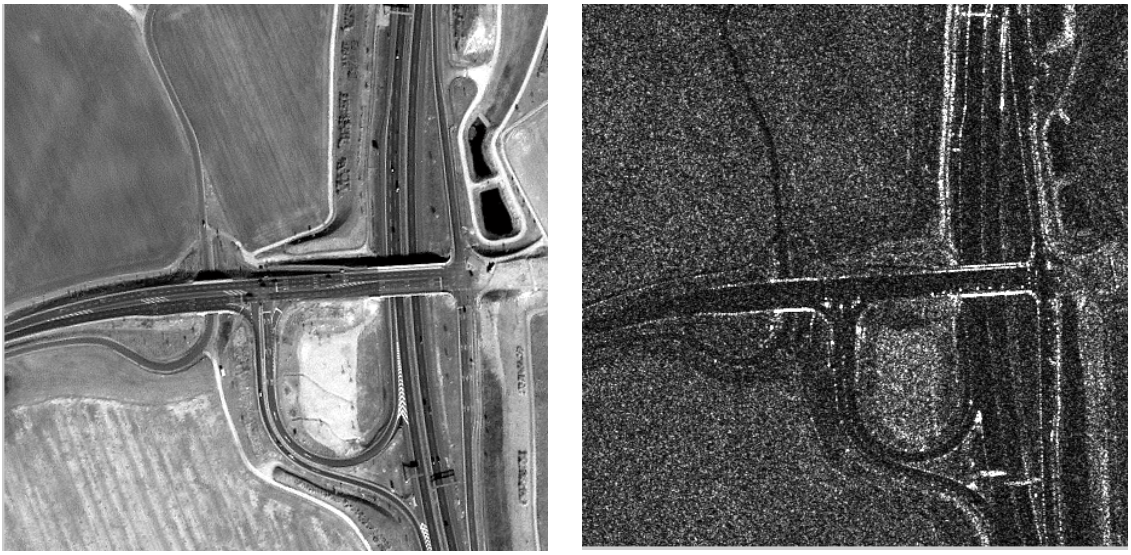
sparse joint histograms generated by remote sensing images, intensity binning or radiometric normalization is recommended. Reducing the joint histogram size would not only help in checking the registration turn around time but has also been found useful to generate smoother MI search spaces (Suri and Reinartz, 2008; Cole-Rhodes et al., 2003). For the presented analysis in this dissertation we have utilized the following linear intensity binning technique

$$DN_{new} = \text{round}\left(\frac{\text{DN}_{\text{org}}}{\text{max\_dn}} \times \left(bin\_size - 1\right)\right) \tag{3.38}$$

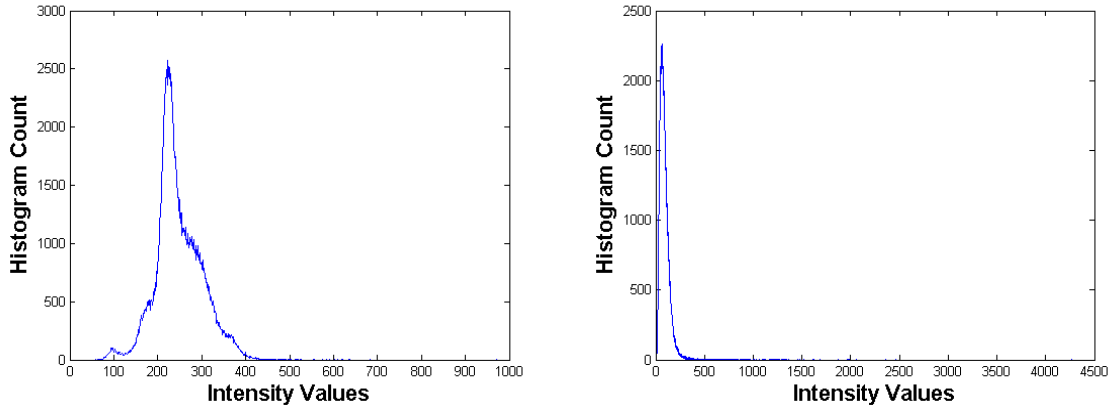
where DN_{org} is the original intensity value, bin_size is the possible number of distinct intensity values in the new range, max_dn is the maximum intensity recorded in the original image and DN_{new} is the newly scaled digital number according to the bin size selected. This above scheme of normalization can be applied after performing a linear stretch of individual image histograms (without any clipping). The reduced sparseness of the joint histogram of the two images due to histogram normalization for bin size 8 (256 grey levels) to 5 (32 grey levels) can be visualized in Figure 3.12. Already cited literature for MI based image registration does not provide any well defined rule governing joint histogram bin size selection. In this regard, Thévenaz and Unser (2000) suggested that for reliable joint histogram estimation, ratio of number of reference image samples to the number of the joint histogram entries should be about 64. We analyze here this recommendation along with other MI similarity metric characteristics for the following joint histogramming techniques using different bin sizes:

i. Nearest neighbour resampling (NN) (two step joint histogramming technique)
ii. Cubic B-spline (order 4) kernel in Equation 3.23 (one step GPVE technique)
iii. Sextic B-spline (order 7) kernel in Equation 3.23 (one step GPVE technique)

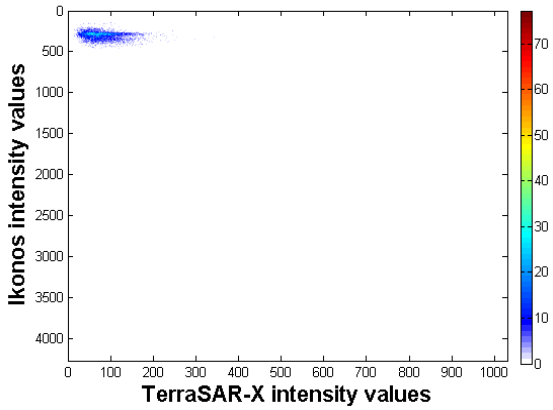
For analysis purposes we select here the pre-registered TerraSAR-X and IKONOS images depicted in Figure 3.11a-b. We translate the input IKONOS image over the reference TerraSAR-X grid in the search space of [-10 +10] pixels in x and y direction at pixel level increments resulting into 441 MI function evaluations for every test case. As only integral pixel translations yielding perfectly aligned grids are considered so in this case nearest neighbour also behaves as a one step joint histogramming technique. Nearest neighbour, Cubic B-Spline and the Sextic B-spline update 1, 9 and 49 bins in the joint histogram respectively (Table 3.1).



(a) (b)



(c) (d)



(e)

Figure 3.11: Sample IKONOS and TerraSAR-X imagery displayed along with their respective histograms.

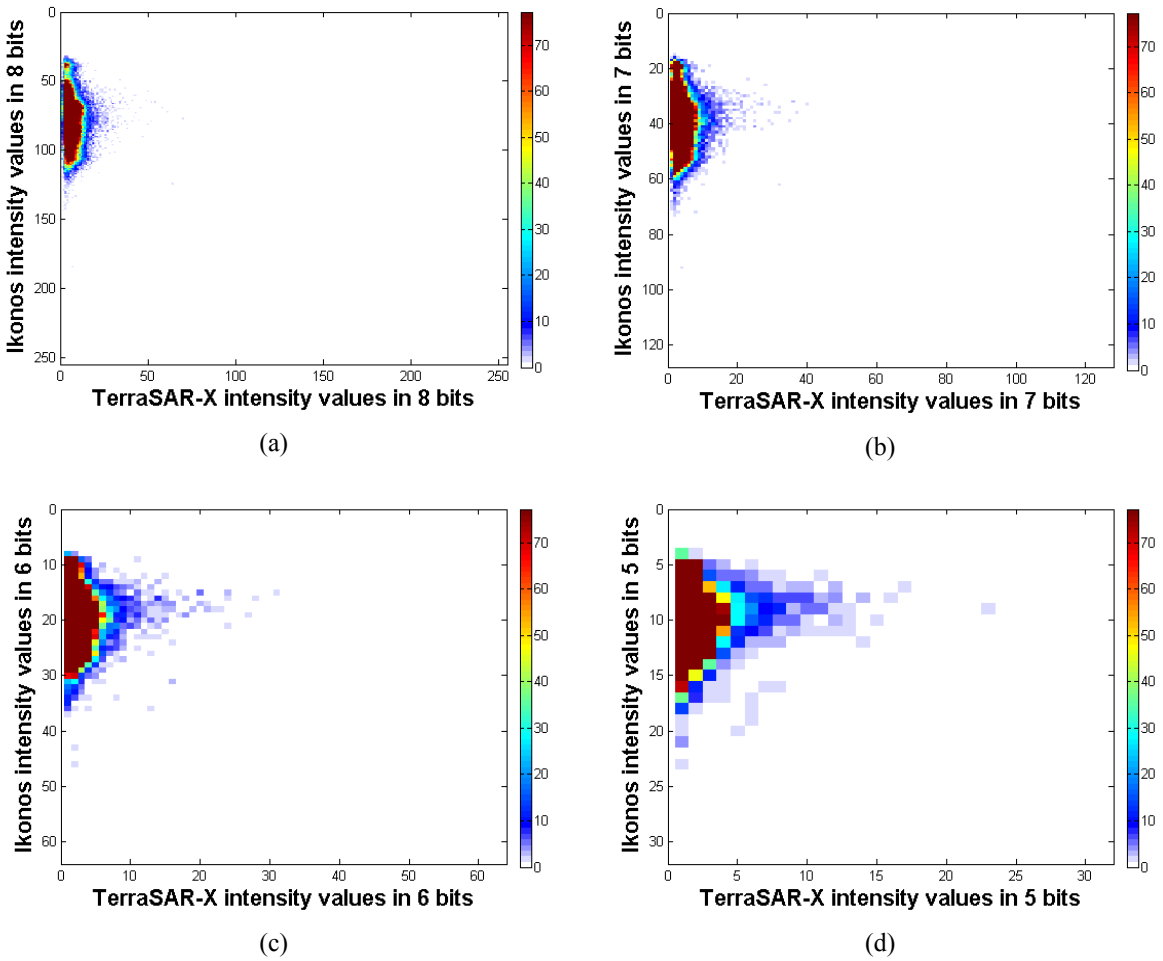


Figure 3.12: Joint histogram of TerraSAR-X and IKONOS image with radiometry of 8-5 bits (a-d).

A glance at the results provided in Table 3.2 clearly indicates a very consistent performance for the Sextic B-spline kernel in the GPVE scheme. Sextic B-splines detect the true peak at (0, 0) consistently for the original image intensity values till to the point where image intensity values have been reduced to 5 bit radiometry (32 bin size). Cubic B-spline kernel shows a small discrepancy for the original image intensity values but then stabilizes till 5 bit image radiometry. The simplest of all the techniques nearest neighbour image resampling shows a little inconsistent behaviour for different bin sizes but it also achieves the assumed true registration peak of (0, 0) for bin sizes 256 and 128.

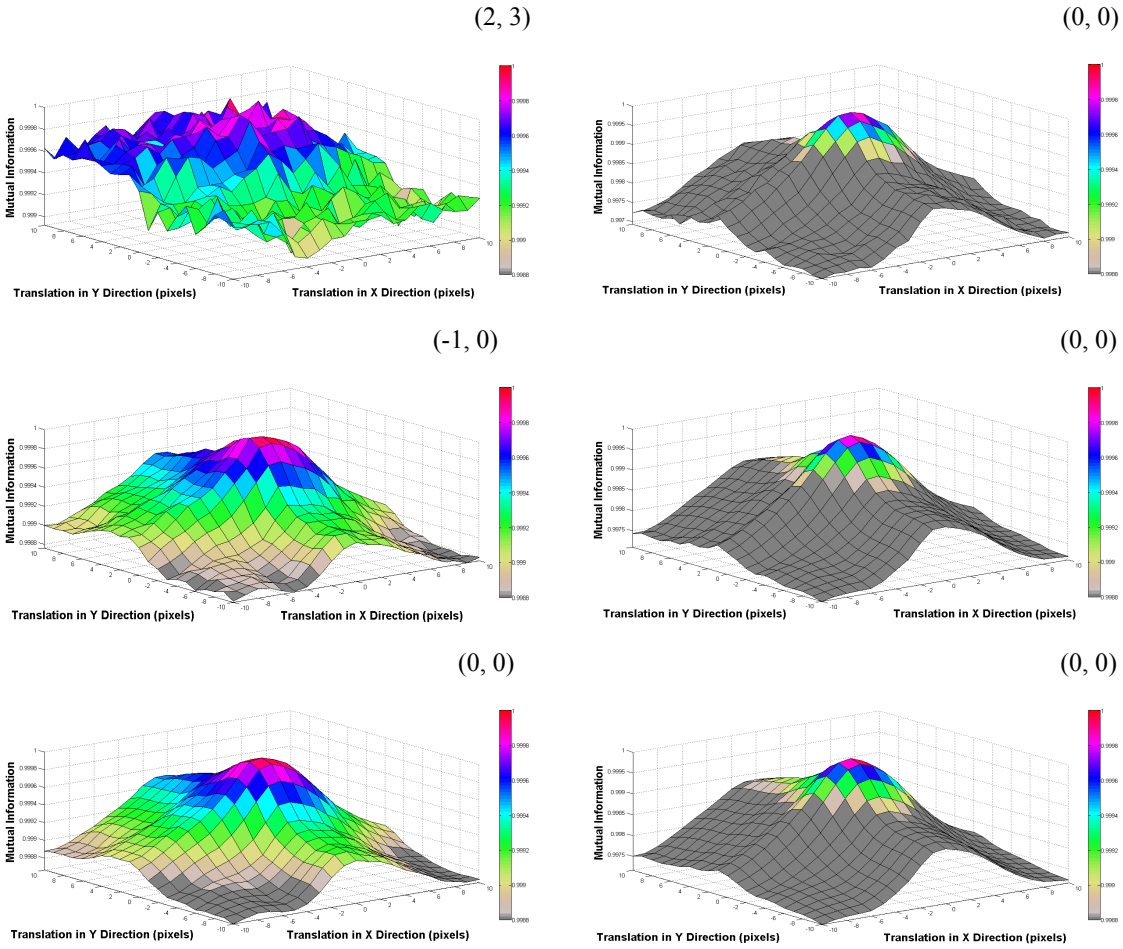


Figure 3.13: Performance of Nearest neighbour (top), Cubic B-spline (middle) and Sextic B-spline (bottom) for original image intensity values (left) and intensity values reduced to 7 bit radiometry (right). The x and y axis represent translation values of [-20, +20] pixels in x and y direction respectively. The z axis represents the normalized MI values obtained. The peak of the registration surfaces are mentioned on the top right

Table 3.2: X and Y co-ordinates of peak produced by MI using different bin sizes and kernels for joint histogram estimation.

Bin Size	Entropy (SAR Image)	Entropy (Opt. Image)	NN	Cubic B-Spline	Sextic B-Spline	JH Entries	Ratio
-	7.40	7.67	(2, 3)	(-1,0)	(0, 0)	472 x 868	0.5
256	3.37	5.66	(0, 0)	(0, 0)	(0, 0)	139 x 130	11.3
128	2.42	4.66	(0, 0)	(0, 0)	(0, 0)	76 x 75	36.0
64	1.59	3.66	(-1, 0)	(0, 0)	(0, 0)	42 x 43	112
32	0.63	2.66	(-1, 0)	(0, 0)	(0, 0)	24 x 25	339
16	0.09	1.76	(-10, 5)	(-10, 10)	(-10, 10)	13 x 13	-
8	0.02	1.07	(-5, 2)	(-8, 5)	(-8, 6)	7 x 7	-

The consistency achieved by the higher order B-spline kernels comes with an incurred cost in terms of the execution time. The sample performance of the nearest neighbour, Cubic B-spline

and Sextic B-spline for original image intensity values and intensity values scaled to 7 bit radiometry can be visualized in Figure 3.13. It can be seen that intensity binning produces much smoother and consistent MI search spaces and thus is expected to facilitate the optimization process. The influence of intensity binning is seen maximum on the nearest neighbour technique where the reduced joint histogram sparseness makes its performance equally compatible with the much complicated higher order B-spline kernels. The question of selecting a suitable or an optimal bin size for MI computations cannot be answered empirically but definitely certain recommendations can be suggested. The suggested ratio of number of reference image samples and the joint histogram entries to be around 64 (Thévenaz and Unser, 2000) has not been found as a limiting condition for this tested dataset and several others tested during this thesis work. MI has shown satisfactory performance for different bin sizes and joint histogramming techniques for a ratio ranging from 0.5 to 339 (Table 3.2). It can be suggested that bin size selection should be done carefully as too low bin sizes can drastically reduce the image entropy contents and hamper the mutual information performance. As observed in Table 3.2, all the three joint histogramming techniques failed to locate the true peaks for bin sizes of 16 and 8 which can be attributed to very less information or entropy content (especially in the SAR image) being left in both the images for MI to deliver reliable results. An important observation to be made from Table 3.2 is the fact that reliable MI performance could be achieved for very low image entropy contents. Specifically for bin size of 32, the SAR image entropy reduced to 8% of its original value (from 7.40 to 0.63) and the optical image entropy reduced to 35% of its original value (from 7.67 to 2.66), still MI could detect correct registration peaks for the B-spline kernels.

The execution times for all the three techniques for different joint histogram bin sizes are tabulated in Table 3.3. The individual execution times for the three histogramming techniques for different joint histogram size (32x32 to 256x256) increase proportionally to the joint histogram size. Therefore, the dependence of MI algorithm on the joint histogram bin size can be categorized as $O(n)$ where n is the joint histogram bin size. Important is the significant effect of the joint histogramming technique on MI computation time can be visualized in Table 3.3, because of the per pixel $O(n^2)$ operation (explained in the previous section) the Sextic B-spline

kernel takes maximum time for producing the most consistent results followed by Cubic B-spline kernel and nearest neighbour interpolation.

Table 3.3: Processing time[7] in seconds for 441 MI evaluations for various joint histogram bin sizes for a reference image size of 451x451 pixels.

Joint Histogram Size	Nearest Neighbour	Cubic B-Spline Order 4	Sextic B-Spline Order 7
4263*1030	143	562	947
256*256	33	447	832
128*128	33	445	823
64*64	33	444	821
32*32	33	444	820

In general, it can be commented that the easiest to implement and fastest in terms of execution time, nearest neighbourhood joint histogramming technique has similar capabilities as shown by higher order B-spline kernels but it is recommended to utilize this scheme only in the scenario where all possible image transformations lead to perfectly aligned grids as in case of non aligned grids the problem of inconsistent behaviour and interpolation artefacts as shown earlier might arise. Further, by utilizing nearest neighbour scheme, the chances of achieving sub pixel accuracy especially for low resolution sensors like Landsat, Radarsat looks bleak but certainly it can be utilized for pixel level registration of high resolution images from TerraSAR-X and IKONOS, and this has been demonstrated in results presented in the following chapters through a proposed discrete optimization scheme for image registration applications.

3.4.4 Search Space and Strategy

After selecting a suitable histogram bin size, joint histogramming technique and the spatial transformation the final task in intensity based registration process is to efficiently estimate the desired registration parameters. Taking into perspective the huge data volume generated by remote sensing sensors, intensity based registration of remote sensing images is computationally expensive. The set of all possible image transformation might just be too big to do an exhaustive searching of the entire set thereby emphasizing the need of an effective search strategy. Search strategy basically employees an optimization technique which tries to maximize the mutual

[7] On an Intel P4 Xeon (3 GHz, 1 MB cache) with 2 GB RAM

information registration function within the images to be registered. As mentioned earlier, optimizers with different mathematical roots have been utilized for this task. In this scenario, the target is to optimize the MI function with respect to the parameters of the decided registration transformation. For example, if the registration deformation between two images is a x and y directional shift then MI function optimization is a simple two dimensional problem but if the registration deformation is modelled as an affine transformation then the optimization task would be a six dimensional problem where six is the number of parameters.

Table 3.4: Details of the utilized Radarsat and Landsat Imagery.

	Radarsat	**Landsat**
Spectral Resolution	C Band (5.3Ghz)	0.52 - 0.90 µm
Pixel Spacing	12.5 m	14.25 m
Bits per pixel	16	8
Scene Size (pixels)	3003x3008	2937x2941
Date of Acquisition	3 July 2007	5 July 2007

However, even after selecting a suitable optimizer, the computation time problem may still persist as MI implementation is heavily dependent upon the image size. To tackle this problem, the use of a multiresolution search strategy is advocated. Multiresolution search strategy to match features using mutual information has been successfully used in Cole-Rhodes (2003a-b) for estimating the rotation and translation parameters within the images. Multiresolution approach presented in the cited reference to register Landsat imagery by maximizing mutual information between the features extracted by Steerable Simoncelli Filters (Simoncelli and Freeman, 1995) is being extended here for more general purpose registration scenarios.

Here we analyze the metric behaviour using the proposed approach to register two large remote sensing images. The dataset here is a pair of orthorectified Landsat and Radarsat imagery with registration differences incorporating scale, rotation and translation in x and y direction, images are visualized in Figure 3.14 and details of the utilized imagery are tabulated in Table 3.4. In Figure 3.14 we also show manually selected 12 check points in both the images which when utilized to built a affine transformation function have a residual of 0.74 pixel (9.25 m) and 0.65 pixel (7 m) in x and y direction respectively. To evaluate the absolute performance of MI based image registration, a set of precise check points from some ancillary source are mandatory, considering no such source is available for the tested dataset we utilize the manually marked

points which have a fair overall consistency but still slight individual matching errors might exist.

The MI performance at different image resolution levels is analysed with respect to these twelve manually marked control points. As extracting conjugate features from multimodal images especially SAR and optical images can be a difficult task so we utilize the entire image information from the two images and use down sampled images to speed up the entire processing. We propose to down sample the images using a simple block mean filter to form individual image pyramids for both the images and then proceed for multiresolution MI optimization using the SPSA algorithm as proposed in Cole-Rhodes et al. (2003a, b). Simple down sampling technique can be utilized effectively for remote sensing images to be registered on basis of intensity.

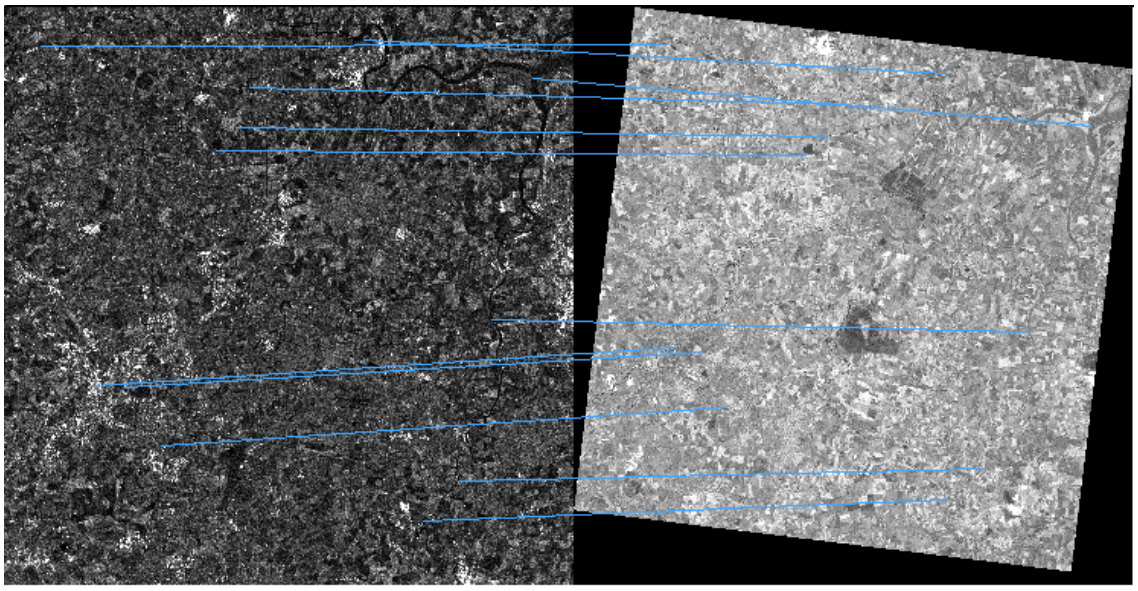
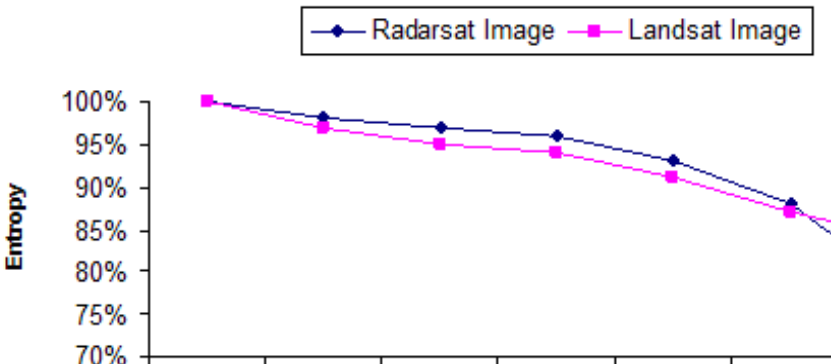


Figure 3.14: Reference Radarsat (left) and the input Landsat (right) image acquired over Yorkshire, United Kingdom in July 2007. The spatial difference between the two images is a rigid body transformation encompassing scale, rotation and translation in x and y direction. The end points of the blue lines represent the tie points utilized for registration accuracy assessment.



	0	4	16	64	256	1024	4096
Radarsat Image	1	0.98	0.97	0.96	0.93	0.88	0.78
Landsat Image	1	0.97	0.95	0.94	0.91	0.87	0.84

Figure 3.15: Impact of image down sampling on image entropy content.

An important argument favouring this is the fact that image down sampling reduces the image size effectively to one-forth at every level does not have a very adverse influence on the image entropy content. This phenomenon is also observed for the images in Figure 3.15 where Radarsat and Landsat images compressed to 6 levels, effectively reducing the image size (64x64) 4096 times still can retain 78 and 84% of their original entropy value (measure of information content). Studying the graph reveals the fact that image entropy is preserved to more than 90% of its original value for both Radarsat and Landsat imagery even when reduced 256 times (level 3) of their original size. This preserved entropy in different levels of the pyramid can definitely improve the registration turn around time for MI based image registration as a lot of information gets represented in far less number of image pixels.

The scaling factor (through image pixel spacing) is normally a known parameter for remote sensing images. Therefore, we optimize the MI registration function with respect to three registration parameters namely, rotation and translation in x and y direction. For the pyramidal MI optimization, a joint histogram bin size of 128 and GPVE with cubic B-spline interpolating kernels has been utilized. The four components of a rigid body transformation (scale, rotation and translation in x and y direction) obtained at different levels of the formed pyramid are provided in Table 3.5. For a metric consistency analysis, we perform a two way optimization of the MI registration function once keeping the Landsat image as the input and Radarsat as the

reference and then the other way around. The metric consistency (RC) is reported using the consistency metric explained in Section 2.5.3.

Table 3.5: Multiresolution MI performance for registration of Landsat and Radarsat imagery.

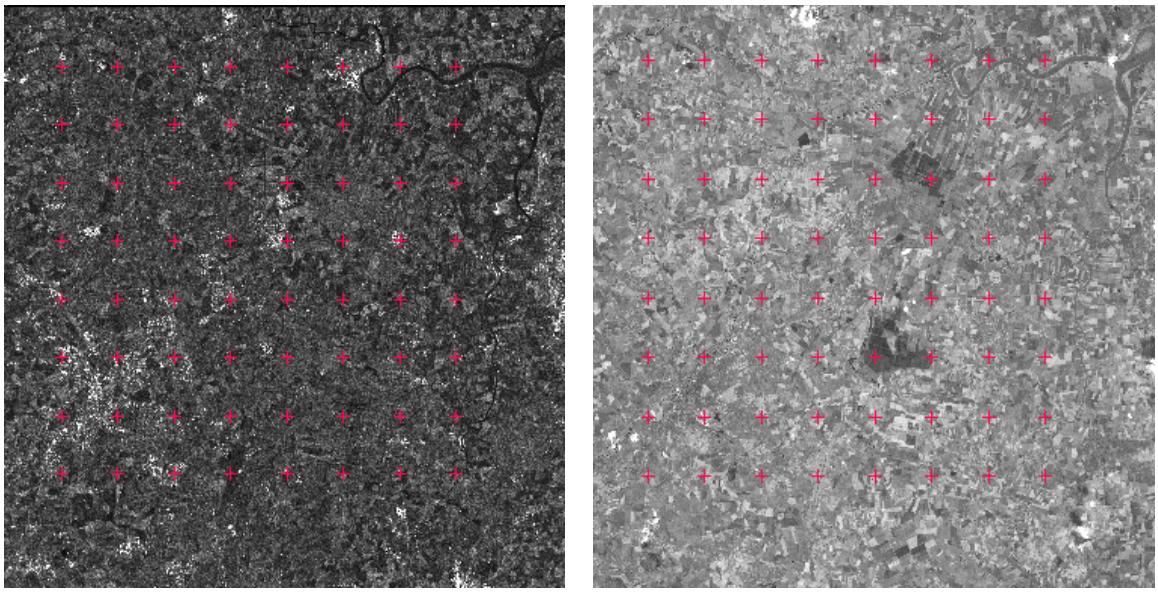
Level No.	$T_{LSAT\ to\ RSAT}$[8]	$T_{RSAT\ to\ LSAT}$	RC (pixels)	FE	TAT[9] (sec)
3	1.14, -6.90°, -12.96, 8.72	0.88, 6.86°, 9.84, -4.4	0.28	336	265
2	1.14, -6.86°, -13.44, 13.92	0.88, 6.74°,10.56, -6.52	0.75	183	600
1	1.14, -6.91°, -14.14, 12.70	0.88, 6.83°, 11.30, -7.60	0.81	114	1440
0	1.14, -6.91°, -14.14, 12.70	0.88, 6.83°, 11.30, -7.60	1.60	54	3060

RC: Registration Consistency, FE: Function Evaluation, TAT: Turn Around Time

The RC value in Table 3.5 is not a measure of the absolute registration accuracy but just depicts the similarity metric performance consistency measured in pixels. We also report the combined number of MI function evaluations (FE) for finishing the optimization task along with the registration turn around time (TAT) at every image pyramid level. It is worthwhile to take a note that fairly similar registration parameters can be estimated from images down sampled to one eight of their original pixel spacing and this can directly be attributed to about 95% image information (entropy) still left in the down sampled images. Further, the TAT parameter value clearly suggests that as the data volume increases, the registration time using MI increase drastically without necessarily improving the registration quality significantly. A limitation of this approach is that registration parameters should be precisely detected by the optimizer at the lowest level of the image pyramid. As the false registration parameters obtained at the lowest level of the image pyramid would propagate to all further layers and the probability of the optimizer to get stuck in a local maximum would increase significantly. Another important factor that might influence the results of this methodology is the optimizer initialization. Every optimizer needs a "seed" value to start the optimization process, in case this value is very far from the true global minimum the probability of the optimizer to halt in a spurious local maximum would increase rapidly. The utilized optimizer in this scenario is SPSA which has been found very effective for rough noisy functions but still this factor can play a significant role on the final results.

[8] The transformation parameters include scale, rotation, x translation (pixels) and y translation (pixels)
[9] On an Intel P4 Xeon (3 GHz, 1 MB cache) with 2 GB RAM

Considering the observed MI performance for estimating the registration parameters, we extend this approach to fine register large remote sensing datasets which might also suffer from fine local differences that generally cannot be taken care of using a simple rigid body transformation. The idea here is simple and can greatly improve registration turn around times for large remote sensing datasets. We simply transform the input Landsat image using the parameters ($T_{LSAT\ to\ RSAT}$) obtained at level 3 of the resolution pyramid in Table 3.5 and then mark a regular grid of tie points in the two images and then refine the tie point pairs through a chip matching technique. To demonstrate this idea and compare its results with the complete multiresolution optimization technique we mark a regular grid of 64 tie points in Radarsat and transformed Landsat image (Figure 3.16). For computing MI function locally around the marked tie points we select a chip size of 300x300 pixels in both the images. Here also we make use of the multiresolution strategy within the local chips and the registration function is optimized for two translation parameters. The shifts obtained for all the 64 tie points in the input Landsat image are provided in Figure 3.17.



(a) (b)

Figure 3.16: Roughly registered Radarsat (a) and Landsat (b) images using the registration parameters obtained from level 3 in the resolution pyramid. A regular grid of 64 tie points is constructed on both the images for local refinement and estimating fine registration parameters.

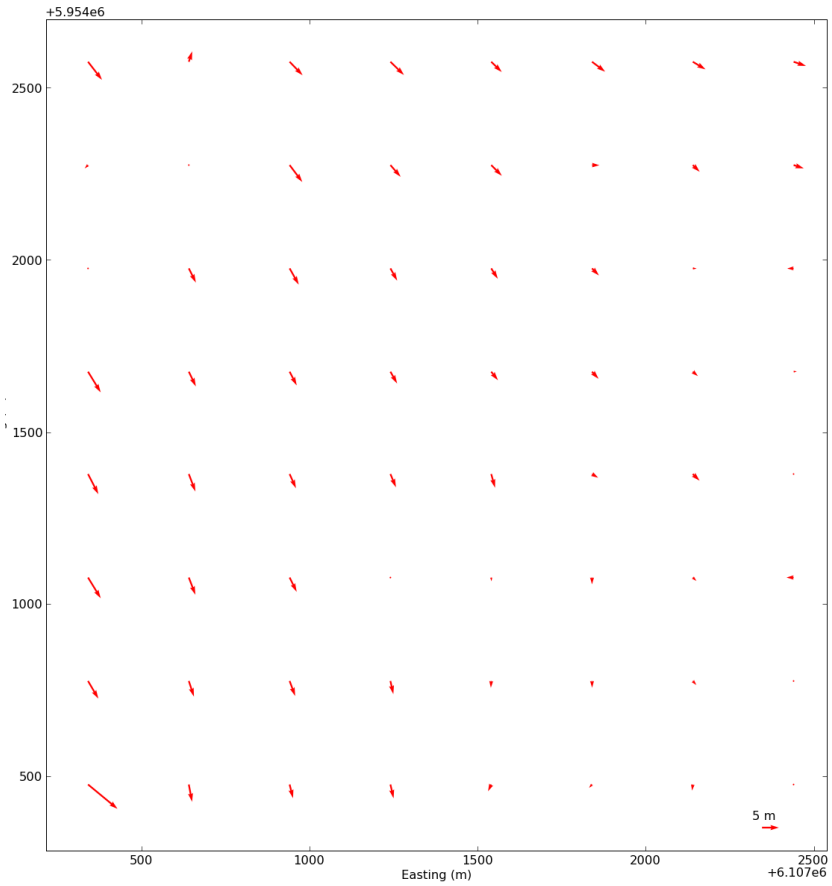


Figure 3.17: Local shifts estimated using local matching using MI for the tie points in the Landsat image (Figure 3.16b).

After the local refinement, possible tie point pair outliers need to be checked for. These outliers might surface due to two main reasons. The first and the most common is the optimizer failure to detect the global maximum and the other possible scenario is lack of enough information in the windows marked around the control points for the metric to produce a sharp enough peak in the registration search space. Both these problems can be taken care of. For the first case optimizer can be initialized differently and change of parameter settings can be done. For the second case the window sizes can be increased to incorporate more image information for similarity metric calculation. In general it is better to use an iterative scheme to detect the outliers. Using the residuals from an initial polynomial built using all the matched points, the most deviating point from the polynomial is removed and a new polynomial is computed. The process is iterated until all returned residuals are smaller than twice the average standard deviation of all the points from the polynomial. Utilizing this scheme on the refined 64 tie point pairs we finally receive 48 tie

point pairs which when utilized to build a first order polynomial transformation function yield a standard deviation of 0.52 pixels in x direction (6.5 m) and 0.41 pixels in y direction (5.12 m). Considering two successive transformations on an image are multiplicative, the final transformation equation between the input Landsat image and the reference Radarsat image is computed as Equation 3.39.

$$\underbrace{\begin{Bmatrix} 1.13 & 0.14 & -18.15 \\ -0.14 & 1.13 & 15.55 \\ 0 & 0 & 1.00 \end{Bmatrix}}_{Final\ transforamtion\ paramters} = \underbrace{\begin{Bmatrix} 1.00 & 0 & -5.19 \\ 0 & 1.00 & 6.80 \\ 0 & 0 & 1.00 \end{Bmatrix}}_{From\ 48\ refined\ conjugate\ pairs} * \underbrace{\begin{Bmatrix} s*\cos(0) & s*\sin(0) & -12.96 \\ -s*\sin(0) & s*\cos(0) & 8.72 \\ 0 & 0 & 1.00 \end{Bmatrix}}_{Level\ 3\ paramters} \tag{3.39}$$

$$where\ s = \frac{Landsat\ pixel\ spacing}{Radarsat\ pixel\ spacing}$$

To visualize the performance of the two different registration schemes, x and y direction deviation in meters between the manually marked control points (after the registration) is plotted in Figure 3.18. From the complete pyramidal optimization scheme, the transformation parameters obtained from level 0 and level 1 were exactly the same in Table 3.6 and thus the results from level 0 are not been shown in Figure 3.18. As expected, the transformed points from the level 3 and level 2 rigid body parameters have huge deviations from the reference x and y axis which slightly improves for the level 1 parameters. The transformed points from the affine parameters (registered Landsat and Radarsat images shown in Figure 3.19) show the minimum deviation from the four transformation parameters but some local differences are observed which might also reflect to inaccuracy in the manually marked control points. This observation is also reflected in obtained shift statistics (Table 3.6) around the twelve manually marked check points shown in Figure 3.14.

Table 3.6: Shift statistics (in meters) around twelve check points at different levels of MI based registration.

	Absolute Average		Standard Deviation	
	X	Y	X	Y
Level 3*	25.42	33.83	20.97	22.57
Level 2*	38.35	23.32	13.9	12.9
Level 1*	19.47	24.99	15.07	16.31
Level 0*	19.47	24.99	15.07	16.31
Affine	9.8	11.32	7.06	8.68

* Rigid body transformation

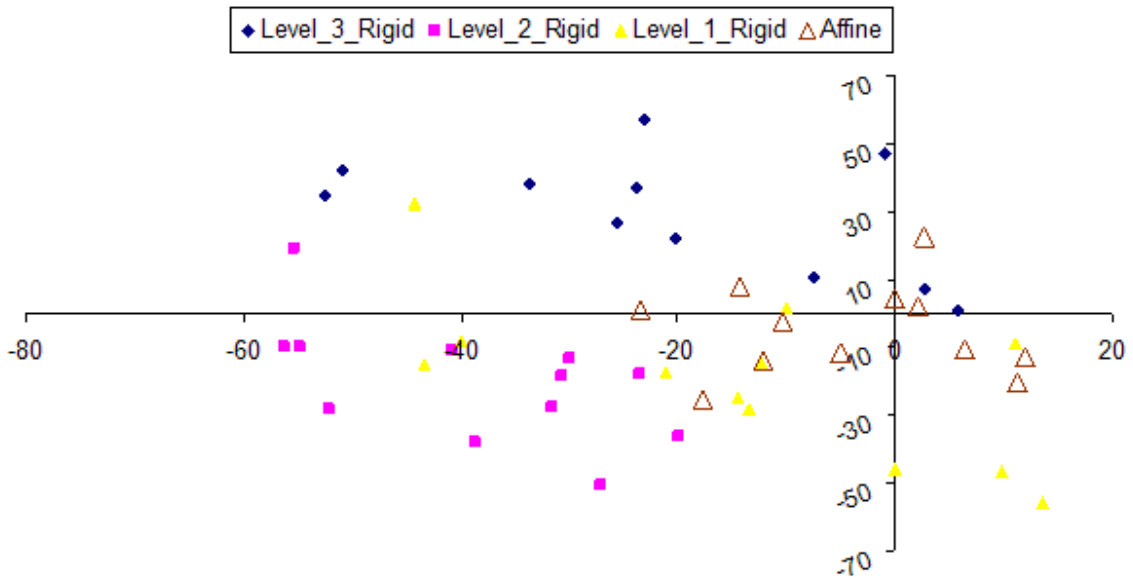


Figure 3.18: Scatter plot of difference in x and y direction within the transformed input image tie points and their manually marked conjugate points in reference Radarsat image. Plot compares the multiresolution optimization scheme (at different levels, Table 3.5) and the estimated affine transformation (Equation 3.39).

In general, the proposed registration scheme has the following advantages for registration of large remote sensing datasets:

Reduced processing time: The behaviour of MI for remote sensing images can be exploited to compress the images for obtaining rough registration parameters thus reducing the computation time extensively.

Adaptability to parallel computing environment: Once the initial rough parameters are estimated from the coarser resolution images each of the conjugate tie point windows can be optimized in a parallel execution environment further reducing the processing time.

Flexibility: The described methodology is flexible to use any number of control point pairs depending upon the kind of accuracy level desired. For certain applications, control points can be deliberately marked in areas expecting some local distortions due to reasons like sensor geometry and terrain influence. The method offers a way to get an estimate of possible local differences according to which a suitable transformation model can be selected.

Model Improvement: This utilized chip matching method can also find immense applications for improvement of sensor model which is investigated later in this dissertation.

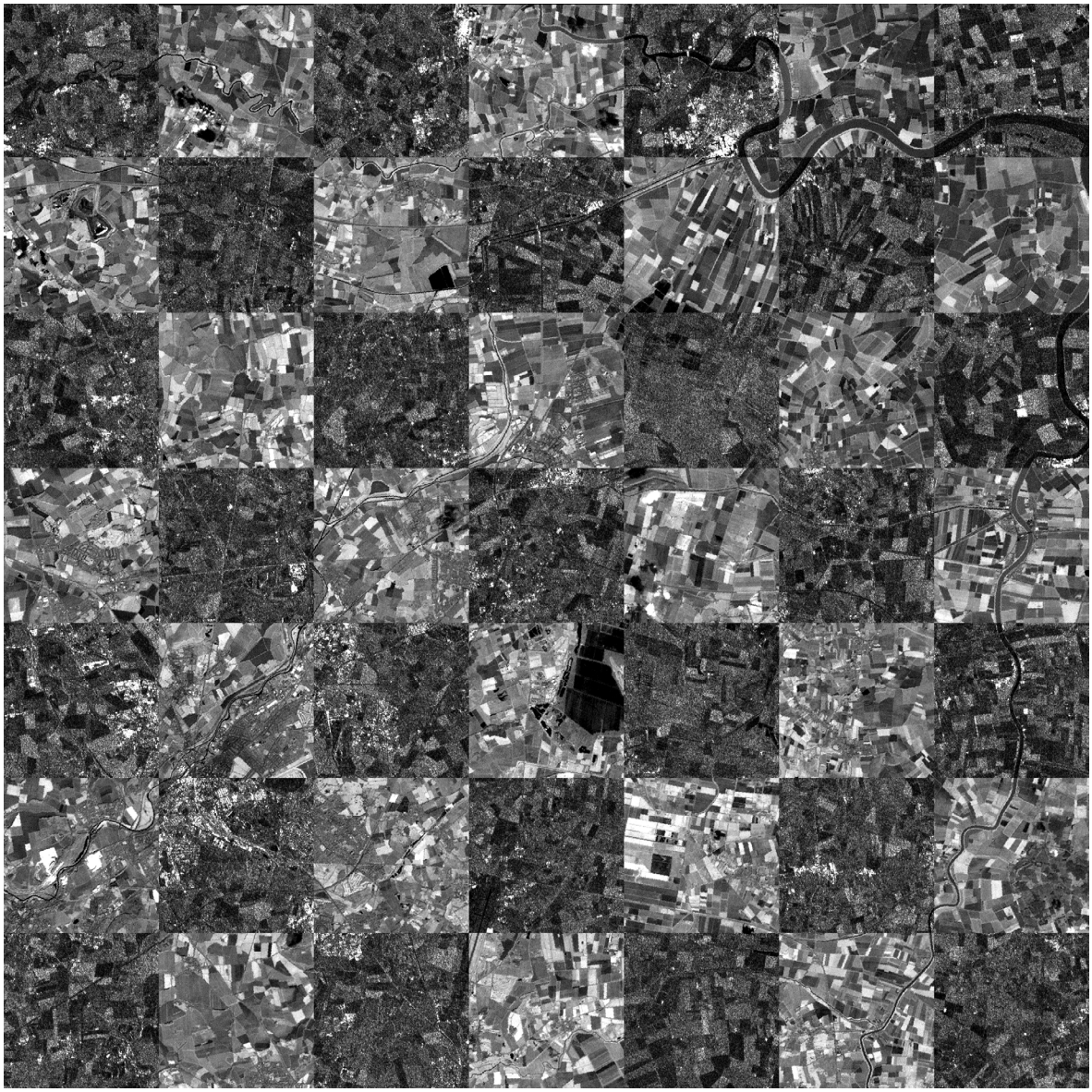


Figure 3.19: Registered Landsat and Radarsat imagery using affine transformation (Equation 3.39).

3.5 CONCLUSION

Considering the meticulous task of extracting similar features from SAR and optical images intensity based methods are preferred. Highlighting the limitations of correlation based methods to handle very different SAR and optical radiometry we have thoroughly discussed the concepts of information theory and formulation of mutual information as a registration similarity metric. The intensity based registration chain utilizing MI has been thoroughly explained with a detailed discussion on very crucial steps of selecting the mapping function, histogram normalization, joint histogramming technique and search space and strategy. A method to efficiently use the similarity metric properties to achieve fine registration for large remote sensing images from medium resolution sensors of Landsat and Radarsat satellites is also suggested. In the following chapter, we discuss performance of MI for images acquired by high resolution satellites where different sensor acquisition principles affect the metric performance adversely.

CHAPTER 4

APPLICATION OF MUTUAL INFORMATION FOR VERY HIGH RESOLUTION IMAGES

4.1 INTRODUCTION

Mutual information has shown enough capability to handle the very different radiometry of SAR and optical images. Compared to images from medium resolution sensors (Radarsat, Landsat, Spot) earth observation data produced by new high resolution satellites like TerraSAR-X and IKONOS are much more affected by different sensor image acquisition principles and also generate huge data volume. In this chapter, we analyse in detail the MI performance for TerraSAR-X and IKONOS images and propose certain steps which might be mandatory for intensity based registration of high resolution images acquired over different land covers.

4.2 UNDERSTANDING HIGH RESOLUTION IMAGES

Considering the fact that MI is a purely intensity based registration method, it is comprehended that significantly incompatible image content and radiometry can influence its performance. With respect to the images of interest, i.e. high resolution SAR and optical in our case, the following concerns need to be acknowledged.

4.2.1 Different Image Acquisition Principles

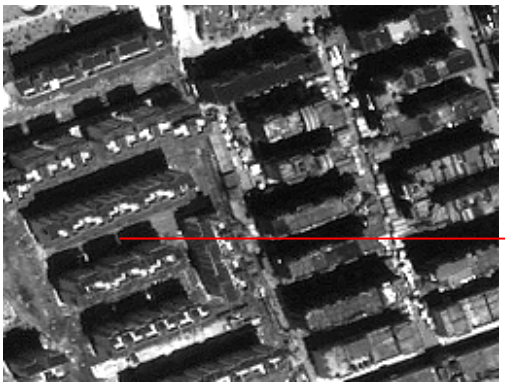
As sensor spatial resolution approaches meter and sub-meter levels, the existing gap of sensor geometry and radiometry between the two sensors further widens up, causing problems for registration methods just based on the raw intensity values (like mutual information). This primarily arises due to the different image acquisition principles (distance versus angle measuring) working in different parts of the electromagnetic spectrum. The SAR side looking

geometry leaves its impact in form of the parallax, shading, unexpected scattering mechanism and double bounce effects especially on 3D world objects (Bamler and Eineder, 2008). Geometric differences leading to very incompatible information becomes clearly visible, especially on man made/natural 3D objects like bridges, buildings, huge towers, and mountains.

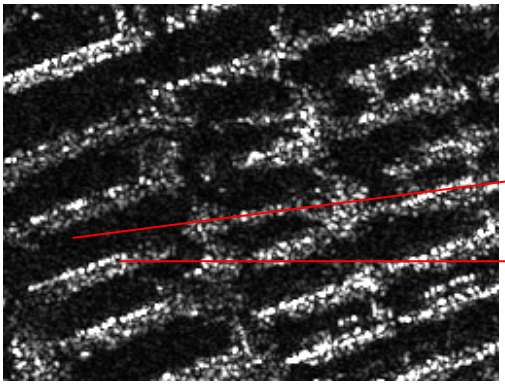
Due to high sensor spatial resolution, all elevated 3D real world objects appear differently (almost incompatible to match) in SAR and optical images. This influence of sensor geometry is not so adverse on plain 2D objects which although appear in different radiometry, but are still not incompatible to match. To better understand this concept of radiometrically and geometrically incompatible 3D objects and only radiometrically incompatible 2D objects an example to depict 3D buildings (Sichuan, China) and a 2D lake (Munich, Germany) as imaged by SAR (TerraSAR-X) and optical (IKONOS) sensors is provided in Figure 4.1. By analyzing the appearance of 3D buildings in urban cities (Figure 4.1a-b) it becomes clear that the object geometry and radiometry are dissimilar in the two images. The sensors produce independent shadow effects in addition to the layover, foreshortening and double bounce effects observed only in the SAR image. On the other hand, the 2D lake provided in Figure 4.2c-d when imaged without occlusions at its border appears independent of any geometry characteristics. In the same scene the influence of different sensor acquisition principle is observed in the 3D trees surrounding the lake. For a detailed understanding of high resolution radar images and the difference from their optical counterparts, interested readers can refer to Stilla (2007). Considering that real world scenes will in generally have a mixture of relatively compatible 2D structures and incompatible 3D structures, the performance of MI for high resolution images comes under a scanner.

4.2.2 Absolute Geometric Accuracy

As the SAR images are a result of a distance measurement principle from space, there accuracy is not much dependent on satellite altitude due to spherical waves which are emitted from the radar antenna. It is due to this reason high resolution SAR especially from the German mission TerraSAR-X shows absolute accuracy in order of 2-3 m for the EEC products which are orthorectified using a DEM, normally derived from SRTM data (Nonaka et al., 2008).

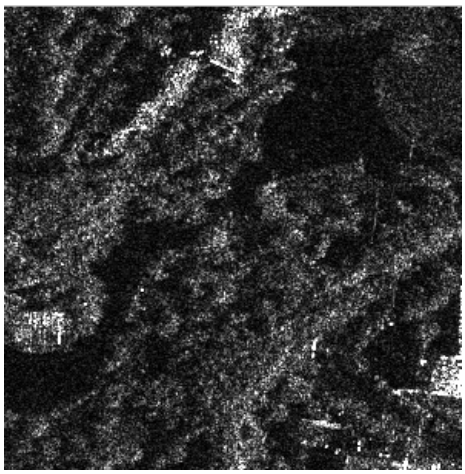


(a)



(b)

(c)



(d)

Figure 4.1: Appearance of real world object in high resolution SAR and optical imagery. 3D buildings (a, b) have a very different response (both radiometric and geometric) where as the 2D lake (c, d) when imaged without occlusions produces much compatible response in both the sensors.

On the other hand, high and very high resolution (VHR) orthorectified optical images using sensor orientation still need ground control information to reach absolute geometric accuracy in pixel or sub-pixel range. This is attributed due to the insufficient knowledge of the sensors altitude and influence of thermal influenced mounting angles. The errors in this case may vary from five to several hundred meters depending upon the satellite configuration. In addition, precise DEM is mandatory for achieving high geometric accuracy (Reinartz et al., 2006b).

4.2.3 Data Volume

The issue of ever increasing data volume has always been and continues to be a matter of concern in remote sensing image processing. As the spatial resolution of modern sensors reach meter and sub meter levels the amount of data that needs to be processed is just huge. In the previous chapter, we highlighted the capability of a simple image down sampling technique to speedup the registration process for Landsat and Radarsat images with approximately 15 m pixel spacing. In general, the MI based registration process was shown to have $O\left(n^2\right)$ dependence on the image size and the computational complexity is further prone to increase with high order B-spline kernels employed in the single step GPVE joint histogramming technique (see Sections 3.4.2 - 3.4.3). These concerns need to be addressed before utilizing MI for automatic registration of huge volumes of high resolution images as now and in near future, processing large volumes of earth observation data is set to become a normal day activity.

4.3 PERFORMANCE EVALUATION IN FLAT PLAIN AREAS

We present here the MI performance for an image pair from TerraSAR-X and IKONOS-2 scenes acquired over west of Munich, Germany. The details of the selected dataset are tabulated in Table 4.1. The utilized TerraSAR-X scene has been orthorectified using a DEM and on the other hand the IKONOS image has been projected to a constant ellipsoidal height during the geocoding process and therefore georeferencing differences exist within the images and these are estimated using MI in this experiment.

Table 4.1: Details of TerraSAR-X and IKONOS-2 Imagery.

	TerraSAR-X	**IKONOS-2**[10]
Mode	High resolution spot light (HS)	Reverse Scanning
Spectral Resolution	9.65 GHz	450 - 900 nm
Spatial Resolution	Range: 0.96m Azimuth: 1.12 m	Cross Scan: 0.83 m Along Scan: 0.82 m
Pixel Spacing	1 m	1 m (panchromatic)
Bits per pixel	16 bit	11 bit
Date of Acquisition	25/12/07	14/10/07
Processing Level	EEC Product	IKONOS Geo

The imagery utilized for this dataset is provided in Figure 4.2, it can be assumed that for these small chip sizes registration differences can be eliminated by estimating shifts in x and y direction.

Figure 4.2: IKONOS-2 (left) and TerraSAR-X (right) imagery. Some of the objects producing incompatible response to the two sensors and should not be participating in any registration process are highlighted.

It is seen in the images (Figure 4.2) that most of the area is covered by plain fields and the road network, which can be put in the category of 2D structures, is not significantly affected by different sensor imaging principles. It is observed that high resolution images have abundant numbers of 'targets' which produce a very strong backscatter especially in the SAR image. Incompatible information in the images being considered are the sign boards on the road highway and poles (marked in Figure 4.2) which have a 3D structure and thus produce very

[10] Courtesy European Space Imaging

different response in the two images. It is the presence of structures of these targets in high resolution images that was much less observed in medium resolution images.

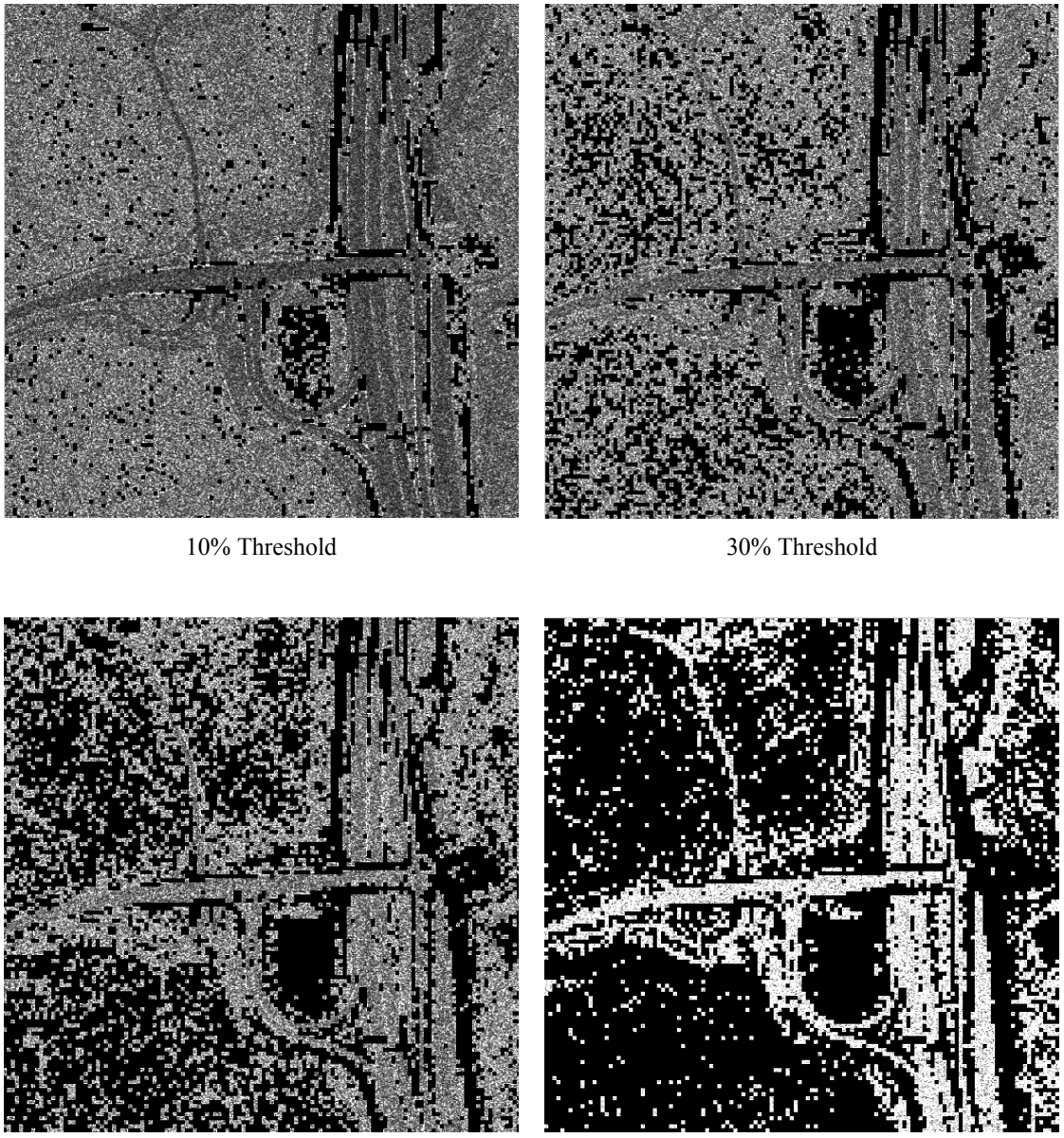


Figure 4.3: Thresholded TerraSAR-X image (Figure 4.2). All the depicted thresholded images produced the same registration peak (Table 4.2) with the corresponding optical image suggesting the fact that complete information from both the images might not be required for robust mutual information performance.

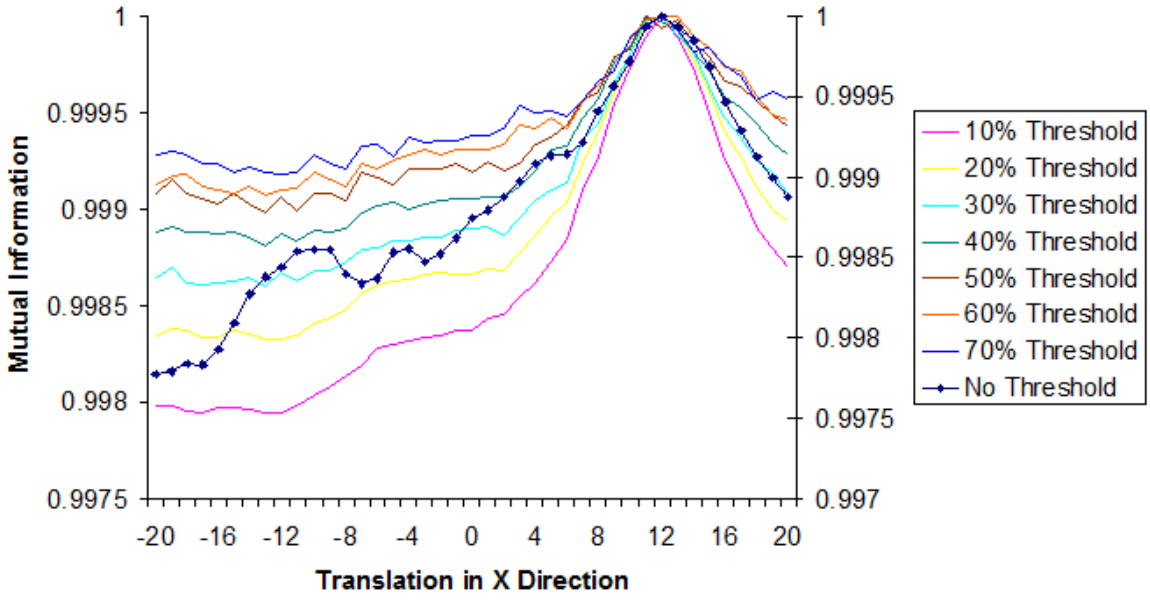
Considering the nature of implementation discussed for MI based registration (see Section 3.4) requiring all image pixels to participate in the registration process, in this case would also

include the incompatible response produced by certain 'targets' now visible in high resolution images. Therefore, the capability and immunity of mutual information to these incompatible targets becomes an important study for high resolution satellite imagery.

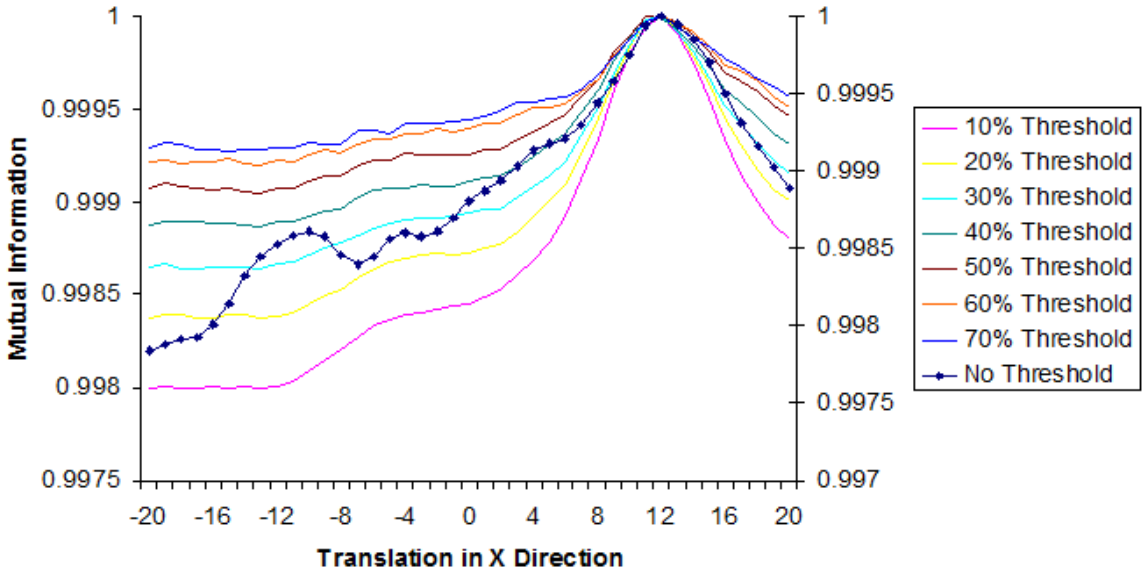
Table 4.2: MI registration peaks obtained between the reference thresholded SAR image and the original optical image. It is observed that suitable feature selection in only one of the images can be utilized to improve the registration processing time considerably without influencing the registration results (Suri and Reinartz, 2010).

Ref Image	Pixel Ref Image (%)	Entropy Ref Image (%)	Peak (K4)	1681 K4 MI evaluations (minutes)	Peak (K0)	1681 K0 MI evaluations (seconds)
Original SAR	100	100	(12, -5)	36	(11, -5)	157
10%TH_SAR	90.0	97.4	(12, -6)	32	(12, -6)	150
20%TH_SAR	80.4	96.4	(12, -6)	29	(12, -6)	136
30%TH_SAR	71.3	95.4	(12, -6)	25	(12, -6)	120
40%TH_SAR	61.2	94.6	(11, -6)	21	(12, -5)	112
50%TH_SAR	51.3	93.6	(12, -5)	18	(12, -5)	84
60%TH_SAR	41.4	92.6	(12, -5)	15	(12, -5)	71
70%TH_SAR	30.2	91.2	(12, -5)	11	(12, -5)	59
80%TH_SAR	20.6	89.7	(12, -5)	8	(13, -8)	38
90%TH_SAR	10.2	87.6	(13, -3)	4	(14, 13)	20
95%TH_SAR	5.6	56.5	(19, 19)	2	(11, -8)	12

A strong high resolution data characteristic is observed in Table 4.2, removing the higher end pixel values from a high resolution SAR image does not influence much of the original image entropy (Shannon) content. It is worthwhile to note that almost 90% of the image entropy is contained in 30% of the image pixels which is utilized to achieve a registration speedup factor of three. This observation can mainly be attributed to the radiometric properties of high resolution SAR sensors where most of the image content towards the higher end of the histogram is redundant (for intensity based registration metrics) mainly generated by real world 3D targets. Abundant number of these targets can actually derail the registration process (shown in subsequent chapter) and therefore should be avoided. With this observation and understanding, selection of fewer relevant pixels from lower end of the SAR image histogram can be taken for an efficient and effective MI based registration.



Nearest Neighbour or GPVE using B-spline kernel (K0)



GPVE using cubic B-spline kernel (K4)

Figure 4.4: Mutual information performance utilizing nearest neighbour (top) and cubic B-spline kernel (bottom) in GPVE technique for input image translated by -6 pixels in y direction moved over the reference grid image from -20 to +20 pixels in x direction. Different thresholded TerraSAR-X images have been utilized as the reference image for the same IKONOS image.

This step can not only improve the processing time dramatically but is an important prerequisite to produce accurate registration results for images from future sensors which are expected to provide much higher resolution (generating more redundant information in the SAR image histogram) earth observation data than the here considered TerraSAR-X and IKONOS sensors. It is seen in Table 4.2, for this case, MI reported same registration peak of (12, -6) (with at max 1 m deviation) in decreasing registration processing time between thresholded SAR (10 to 60%) and the optical image, strongly indicating the fact that this kind of a smart feature selection step can be done for fast and robust mutual information performance.

Another key point to be observed is the performance of cubic B-spline (K4) and nearest neighbour (K0) both being one step joint histogramming techniques produced very similar MI peaks in the tested search space. Further, as also depicted in the previous chapter the nearest neighbour technique has again shown its capacity to detect correct registration parameters when we always have perfectly aligned reference and transformed input image grid. The performance of the two joint histogramming techniques for thresholded SAR image and the optical image can be visualized in Figure 4.4. For plots in Figure 4.4 the reference input image has been translated -6 pixels in y direction and then has been shifted from -20 to +20 pixels over the reference image grid which is basically the SAR image thresholded to different levels (0 to 70%). It is observed that nearest neighbour technique produces very similar MI peaks (Figure 4.4) with a speed up factor of around 10 (Table 4.2) when utilized for perfectly aligned input and reference grids. A vigilant observation of the graphs presented in Figure 4.4 also indicates that the registration peak quality declines (for both K0 and K4 kernel) as the thresholding levels in the TerraSAR-X image is increased successively from 10 to 70%. This trend indicates the adverse impact of loss of image information on mutual information performance. In applications involving MI for high resolution imagery, this trade off between quality of registration peak and reduced processing time needs to be balanced for fruitful results.

4.4 DISCRETE OPTIMIZATION FOR IMAGE REGISTRATION

Till date, efforts have been concentrated on developing robust joint histogramming techniques to achieve smooth MI search spaces to facilitate the optimization process in obtaining the desired registration parameters. In general, the joint histogramming techniques which achieve this objective are computationally expensive as these involve smoothing in the joint histogram space. Due to ever increasing data volume, computational dependence of the considered metric on image size and underlying joint histogramming technique, increasing registration processing time is a persistent problem. Considering the nature of high resolution images and encouraging performance of simple but fast nearest neighbour technique a discrete optimization technique (based only on perfectly aligned grids) for high resolution images can be developed.

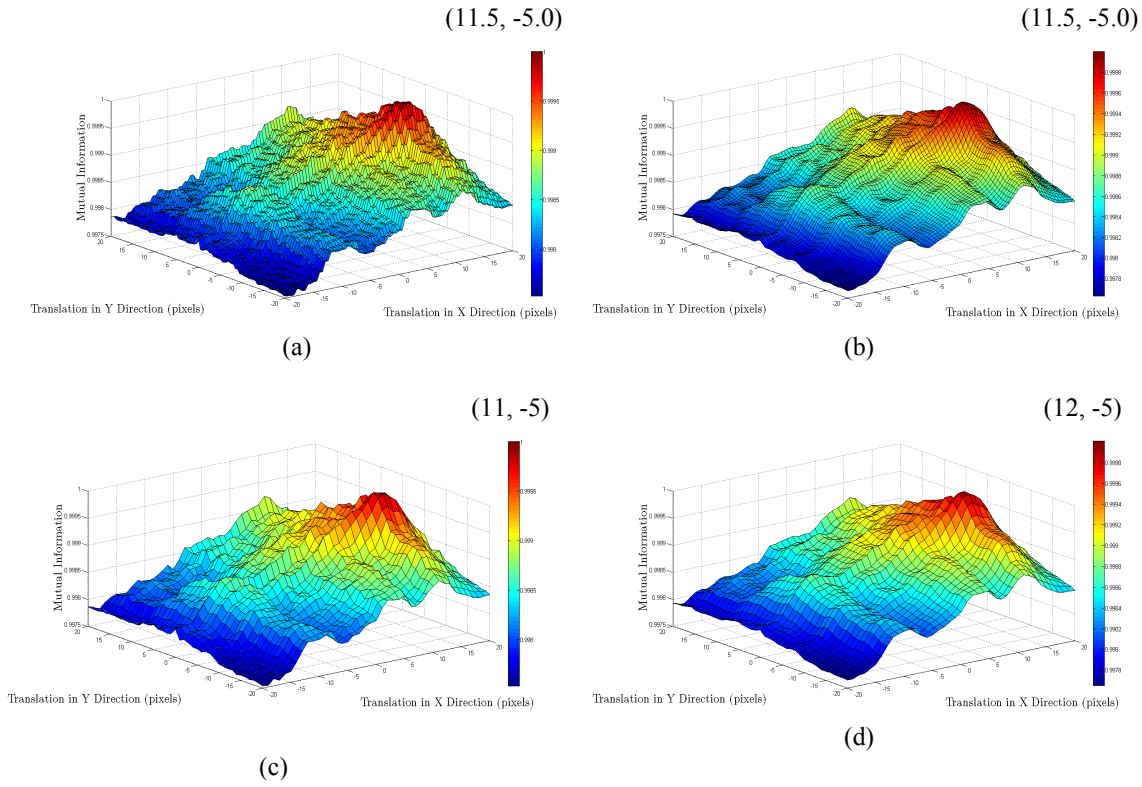


Figure 4.5: Mutual information search space generated by nearest neighbour technique using with (a) subpixel and (c) pixel shift in x and y direction. Mutual information search spaces generated by cubic B-spline kernel with (b) subpixel and (d) pixel shift in x and y direction. It is seen that nearest neighbour produces similar MI surface as generated by cubic B-spline kernel in case of pixel level shifts. The x and y axis represent translation values of [-20, +20] pixels in x and y direction respectively. The z axis represents the normalized MI values obtained. The peak of the registration surfaces are mentioned on the top right

The concept of developing a discrete optimization scheme can be realized by observing the search spaces generated by MI for images of the dataset analysed in the previous section. As before, we translate the input IKONOS image at sub pixel and pixel level in search space of [-20 +20] pixels over the reference TerraSAR-X image. We employ two joint histogramming techniques namely nearest neighbour and GPVE using cubic B-spline to generate search MI search spaces in Figure 4.5. It is clearly observed in Figure 4.5b that the computationally expensive cubic B-spline generates a much smoother search space as compared to nearest neighbour technique (Figure 4.5a) on utilizing half a pixel increment in image translation. On the contrary, search spaces generated using pixel level increments are comparatively similar (Figure 4.5 c, d). This observation can be utilized for improving the registration turnaround time by employing nearest neighbour technique and a discrete optimization scheme which optimizes the MI function solely based only on the values obtained by perfectly aligned reference and transformed input image.

To achieve this, we need to incorporate an optimizer which only works with discontinuous estimates of the mutual information registration function. The optimizer that we select here for the proposed adaptation is the Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm. In its original proposed form, it was also used in the previous chapter to handle registration of Landsat and Radarsat imagery (see Section 3.4.4) through a multiresolution optimization scenario with cubic B-spline kernel being utilized to estimate the registration function in a continuous search space.

4.4.1 Basic Simultaneous Perturbation Stochastic Approximation Algorithm

The SPSA or Simultaneous Perturbation Stochastic Approximation algorithm was originally proposed by Spall (1992). Taking the task of optimizing a real valued function $f(\theta)$, defined on an open domain D in $p-$dimensional Euclidean space R^p. The function $f(\theta)$ is assumed to be at least three times differentiable and have a unique minimum point in D. It is assumed that measurements of $f(\theta)$ over different points in D are available. As the algorithm tries to

optimize the function based on its gradients, the SPSA uses two measurements of f at an iteration k to form a gradient estimate

$$M_k^+(\theta) = f(\theta + c\Delta_k) \tag{4.1}$$

$$M_k^-(\theta) = f(\theta - c\Delta_k) \tag{4.2}$$

The SPSA gradient estimate of the gradient at k^{th} iteration is

$$g(k,\theta) = \frac{\left(M_k^+(\theta) - M_k^+(\theta)\right)}{2c\Delta_k} \tag{4.3}$$

$$\theta_{k+1} = \theta_k - a_k g_k(\theta_k) \tag{4.4}$$

Where c and a are positive valued gain parameters, and Δ_k is an independent random Bernoulli sequence of ± 1 generated for all the function arguments. In the considered example it is assumed that the value of function f is dependent only upon a single variable θ, which is not the general case. The gradient is estimated using the above three equations for all the parameters of the function being optimized. The basic algorithm can be suitably modified to

i. Maximize or minimize the target function *registration*
ii. Generalize for a *n* argument target function
iii. To decide exit conditions prior to the set number of optimizer iterations
iv. Further, to do a constraint optimization with pre-defined search domains for the function parameters being optimized.

The SPSA algorithm is highly parametric and the choice of the gain parameters especially (a, c) is critical to the performance of SPSA (Spall, 1998). Here we utilize the same values for parameters as Cole-Rhodes et al. (2003a-b) utilized for both first and second order SPSA optimization. Utilizing the recommended optimizer coefficients, we test the optimizer performance by maximizing the mutual information registration function to estimate the shift parameters (x and y direction) within the TerraSAR-X and IKONOS images of dataset 1 (Figure 4.2).

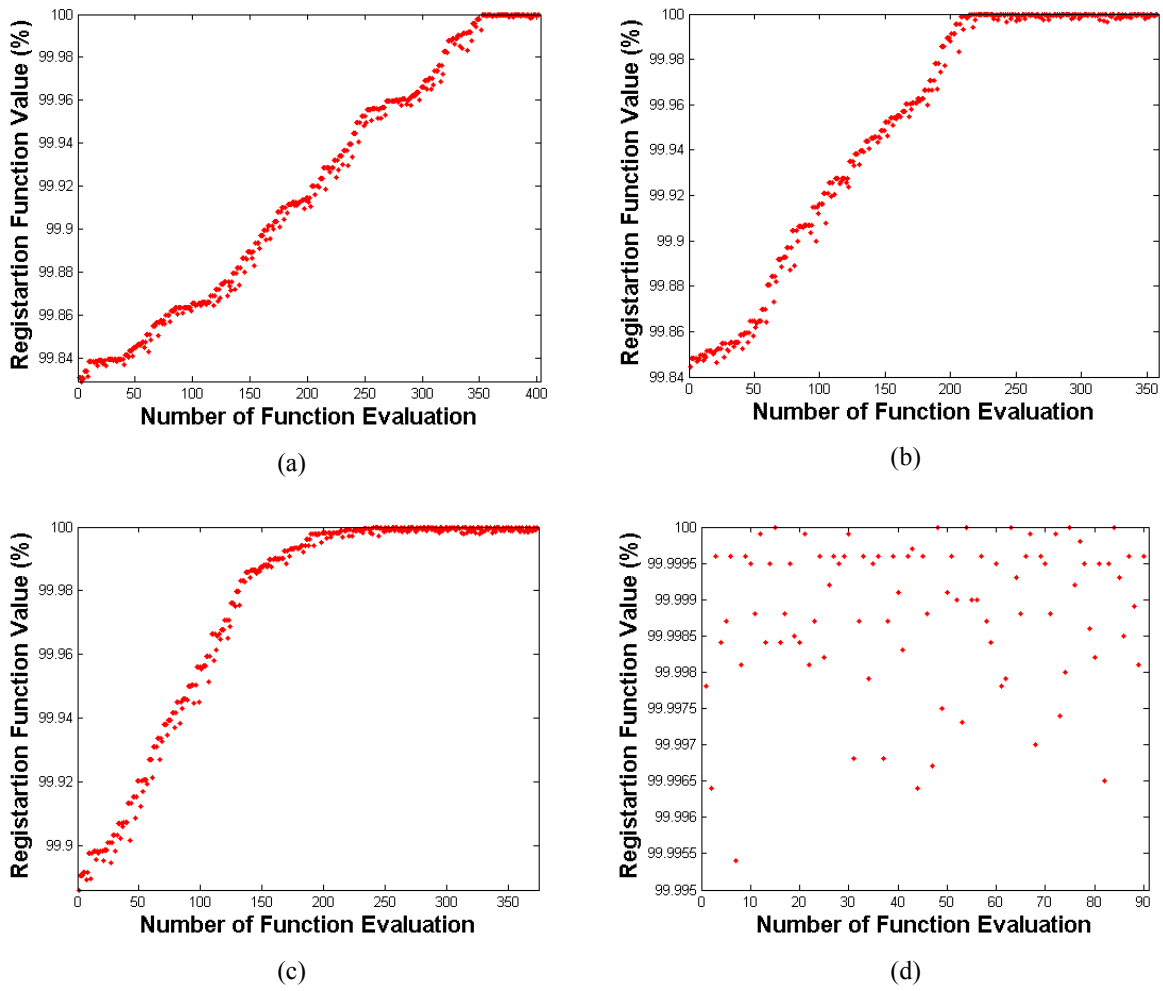


Figure 4.6: Maximization of the mutual information registration function to estimate the shifts between the TerraSAR-X and IKONOS imagery shown in Figure 4.2. (a-b) Represent successful optimization of the registration function using the SPSA algorithm with initial seed being (0, 0) and (20, -2) respectively. (c) The optimizer converges to a false maximum with seed (-20, 2). (d) The optimizer is stuck in a local minima when initialized with seed (-11, 5) respectively.

Observing the MI search spaces presented in Figure 4.5, the shift parameters are expected to be around (11, -5) while keeping the more accurate TerraSAR-X as the reference and the IKONOS image as the input image. Generally all the optimization algorithms need a seed initialization to kick start the optimization process. We present the performance of the basic SPSA for four different initializations in Figure 4.6a-d. Figure 4.6a-b represent the successful optimization of the registration function with shift parameters being (11.55, -5.11) and (11.27, -5.00) for optimizer seeds (0, 0) and (20, -2) respectively which are very near to the expected maximum of (11, -5). Figure 4.6c represents the case where optimizer converges to a false maximum at

(12.04, 0.12) when initialized by (-20, 2). Figure 4.6d represents the case where the optimizer terminated at the initialization seed of (-11, 5) as the registration function value did not change significantly for 40 successive function evaluations which has been set as a terminating criteria for the optimizer.

Considering the nature of registration search spaces generated by MI (especially incorporating misaligned grids, see Figure 4.5b) even when higher order kernels are employed for joint histogram estimation, the optimizer can halt at a false spurious peak. This remains a concern in utilizing MI for multimodal registration. One normally utilized solution is to double check the registration peak by performing a two way optimization by interchanging the reference and input image and evaluating registration metric consistency (see Section 2.4.3) to be doubly sure of the obtained maximum. This method comes with an additional cost of a two way optimization and is obviously not full proof as it fails to detect registration failure in case the optimizer halts at the same false peak in both optimization attempts.

4.4.2 Discrete Simultaneous Perturbation Stochastic Approximation Algorithm

In this sub section, we propose to utilize a discrete optimization scheme. However, it needs to be mentioned that this discrete optimization can only be utilized when we have only a translational shift difference between images having the same pixel spacing. Generally, remote sensing images have well known ground sampling distance and therefore images (small chips) can be brought to the same pixel spacing using appropriate resampling techniques. For geocoded imagery, rotational differences seldom exist within fractions of images, therefore discrete optimization being introduced here is expected to find applications especially for high resolution imagery where nearest neighbour technique has similar accuracy as compared to higher order B-spline kernels.

The already explained SPSA optimizer is suitably modified to achieve the desired target. Another sample example of discrete optimization of SPSA algorithm for the problem of constrained discrete resource allocation is found in Hill (2005). In the literature reviewed for this dissertation, attempt of utilizing discrete optimization in the field of image registration is found to be missing.

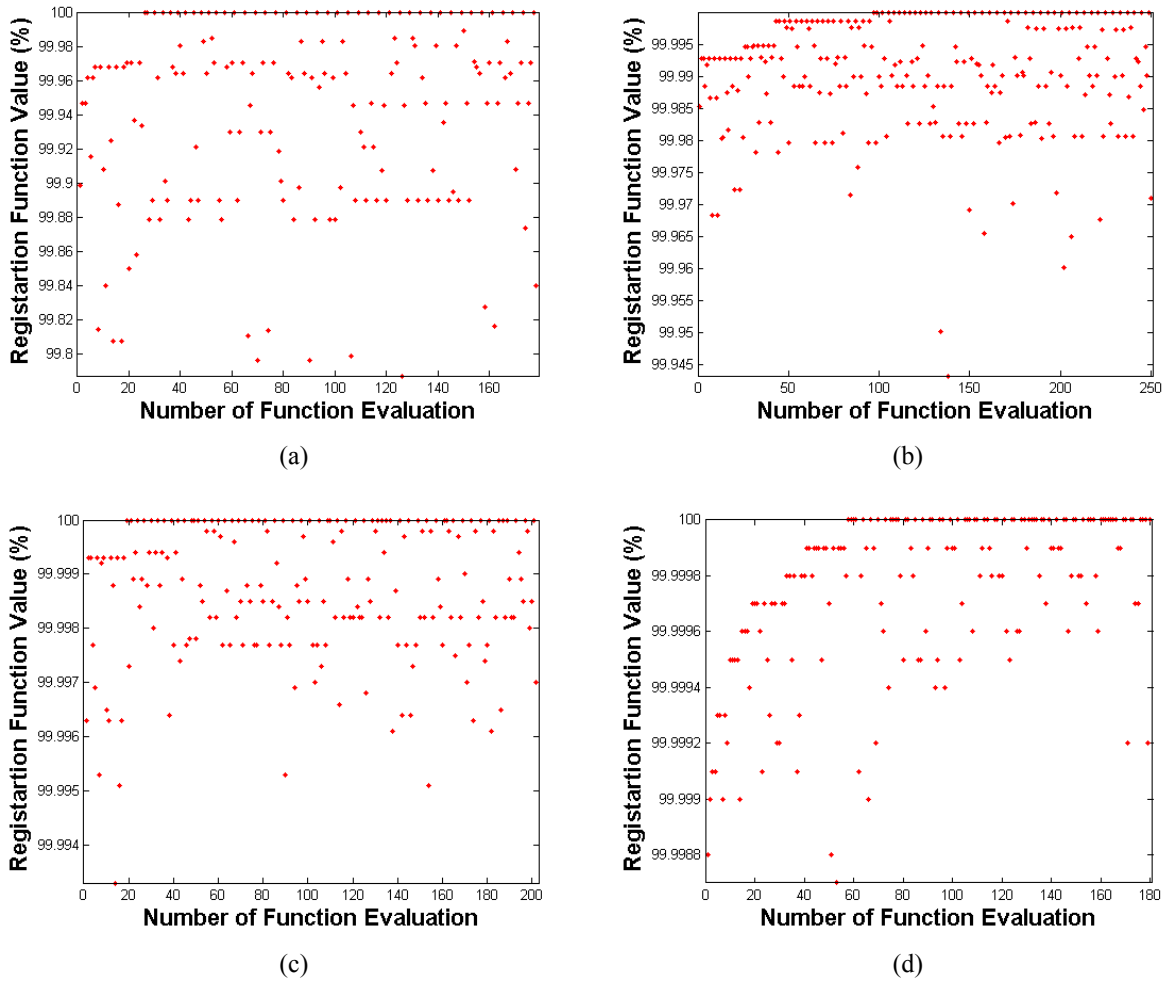


Figure 4.7: Discrete maximization of the mutual information registration function to estimate the shifts between the TerraSAR-X and IKONOS imagery shown in Figure 4.2. Here, the discrete optimization scheme converged to same maximum (-11, 5) with initial seeds being (0, 0), (20, -2), (20, 2) and (-11, 5) for (a-d) respectively. Due to the discrete estimates and increased value of the step coefficient 'c', the optimizer jumps more abruptly here as compared to the continuous case in Figure 4.6.

We use the same parameter setting as utilized in the previous section and slightly modify the optimizer gain coefficients to achieve successive discrete estimates of the MI registration function. Another simple change is to insert round functions in Equations 4.3 and 4.4 respectively to always achieve discrete optimizer candidates which are the two translation parameters in the case of image registration.

The idea behind introducing the round function is to always introduce integral updates in the function arguments of the *registration* function being optimized. We present the performance of

the discrete SPSA using the same initializations of (0, 0), (20, -2), (20, 2) and (-11, 5) as done in the previous section. The optimizer behaviour for these initializations can be observed in Figure 4.7. It is observed that the discrete optimizer using the nearest neighbour interpolation could achieve the true maximum of (11, -5) but this might not be true for certain different optimizer initializations. The reader has to be cautious while interpreting the results as these necessarily doesn't imply that the new optimizer coefficient settings are better than the previous ones. Due to the discrete translation and increased value of the step coefficient 'c', it is expected to see the optimizer jump more abruptly in Figure 4.7 as compared to the continuous case in Figure 4.6.

Table 4.3: SPSA performance[11] for MI based registration (reference Figure 4.6 and 4.7).

	SPSA using Cubic B-spline kernel			Discrete SPSA using nearest neighbour interpolation		
Seed	***Optimizer Iterations***	***TAT***	***Result***	***Optimizer Iterations***	***TAT***	***Result***
(0, 0)[a]	101	446 sec	Optimized to (11.55, -5.11)	65	17 sec	Optimized to (11, -5)
(20, -2)[b]	95	420 sec	Optimized to (11.27, -5.00)	77	22 sec	Optimized to (11, -5)
(-11, 5)[c]	29	143 sec	Optimization failed	118	33 sec	Optimized to (11, -5)
(20, 2)[d]	101	446 sec	Optimization failed	64	17 sec	Optimized to (11, -5)

The chief advantage of the discrete scheme is observed from the statistics in Table 4.3 where the number of optimizer iterations and execution times for the two scenarios are provided. Due to the underlying very simple joint histogramming technique the discrete SPSA achieves the registration peak in far less time as compared to the normally utilized SPSA technique. To achieve further a sub pixel registration result the output of the discrete SPSA can be utilized as a seed to facilitate faster convergence of the normal SPSA using sophisticated joint histogramming techniques.

[11] Experiments done on Genuine Intel Pentium D CPU (2.8 GHz) with 2 GB RAM

4.5 PERFORMANCE EVALUATION IN SEMI URBAN AREAS

It is observed that simple image down sampling and slight modifications in the optimization process can be utilized to enhance the registration time and effectively handle large data volume generated by high resolution sensors. We now analyze the influence of very different image acquisition principles on the performance of mutual information registration similarity metric. Here, we perform our analysis with high resolution imagery acquired over sub urban area in west of Munich, Germany (a subset extracted from the same complete scene utilized for dataset-1). The images can be visualized in Figure 4.8a-b, the imaged scene has urban settlement situated very next to vast plain fields providing an opportunity to analyze the similarity metric performance for two different land covers (both independent and combined).

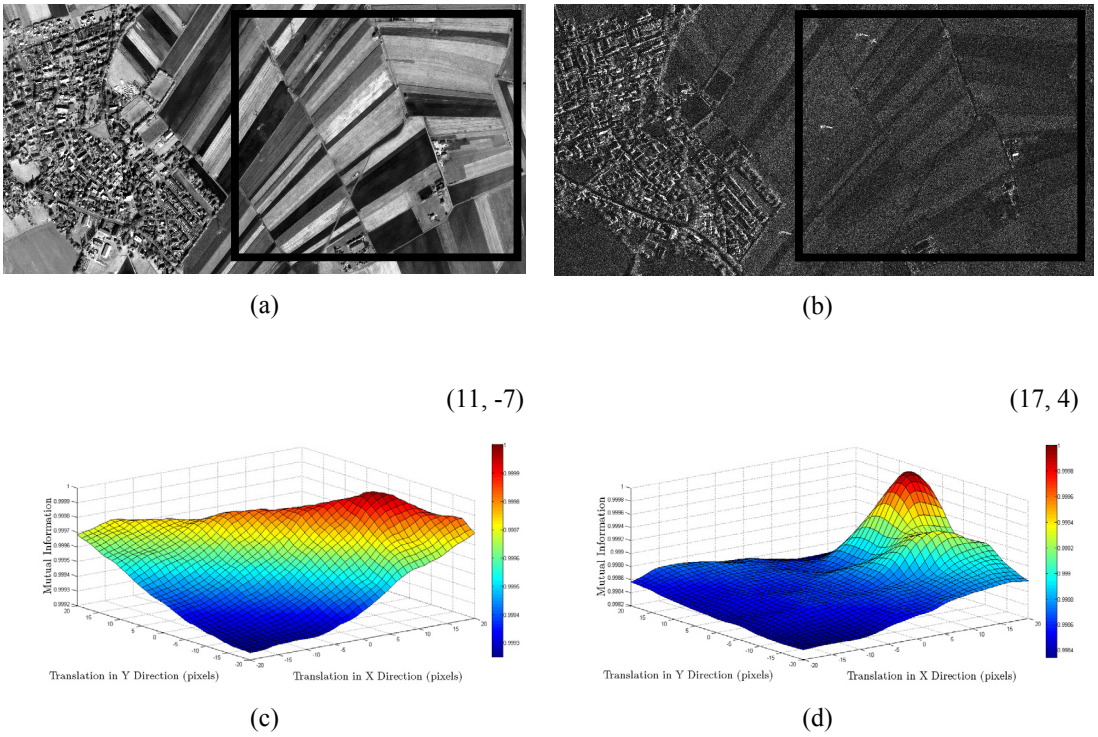


Figure 4.8: IKONOS (a) and TerraSAR-X (b) imagery acquired over semi urban areas. MI statistics estimated using only plain field pixels lead to a registration peak of (11, -7). (c) Introduction of urban area pixels in MI statistics shifts the registration peak to (17, 4) (d). The x and y axis represent translation values of [-20, +20] pixels in x and y direction respectively. The z axis represents the normalized MI values obtained. The peak of the registration surfaces are mentioned on the top right

The following two scenarios have been considered:

i. For case 1, we select pixels only from the plain fields (roughly demarcated with rectangle in Figure 4.8) in both the images for computing the registration parameters (Size: 953x1096 pixels). For the plain field pixels, the side ways looking SAR sensor and the downward looking optical sensor are not expected to have much of their geometric influence so favourable registration results as in the previous case are expected.

ii. For case 2, we select the entire image scene for registration parameter computation and hence analyze the influence of the sub urban establishments on similarity metric performance (Size: 1001 x 2001 pixels). The urban establishments due to their strong radar backscatter cause great changes in SAR image radiometry (histogram information) and thus offer a testing case for the similarity metric.

The search spaces generated in Figure 4.8c-d represent the MI response obtained by moving the input IKONOS image over the reference image grid in a predefined range of [-20 20] pixels in both x and y direction. Figure 4.8c and 4.8d represent the generated search spaces for MI while utilizing pixels belonging only to the land cover class fields and the complete image region respectively. For the two cases, MI reported a peak at (11, -7) and (17, 4) respectively. A visual analysis using an overlay tool clearly indicates the present misalignment within the imagery after using the obtained registration parameters from case 2. Although the land cover fields constitute more than 65% of the total image area but still the introduction of the bright urban area pixels have derailed the registration process. This shift in the registration peak can directly be related to the introduction of region generating incompatible radiometric information (3D real world objects) which has noticeably hampered the mutual information performance for achieving on ground co-registration. The magnitude and direction of this observed shift in the MI peak (which to evaluate quantitatively might require detailed scene information like a DSM and sensor acquisition parameters and settings) in actual is dependent on one or more of the following factors:

i. Sun elevation angle (influences response of urban 3D objects in optical images)

ii. SAR incidence angle (influences response of urban 3D objects in SAR images)
iii. Building shape, height and their orientation towards the SAR sensor (influences double bounce pattern and radar shadow observed in SAR images acquired over urban areas)

Practically, different land cover classes are hardly as segregated as available in the analysed dataset. Therefore, the problem of mixed land cover classes asks for a segmentation step before intensity based registration technique. The segmentation should principally be targeted to incorporate only those pixels in the registration process which are less influenced by different sensor geometries (like the plain field pixels in dataset 1 and 2).

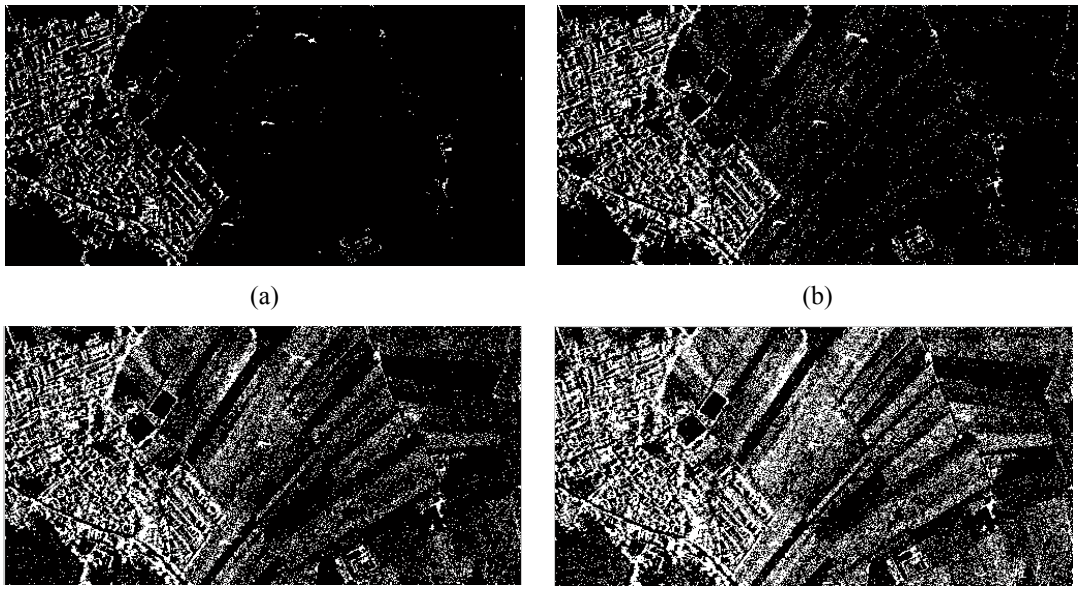


Figure 4.9: Pixels with value 1 were left out (in SAR image) of the registration process after introducing high thresholds of 5% (a) 10% (b) 20% (c) and 30% (d) at image compressed to one-forth of its original resolution.

However, the idea of introducing a segmentation step prior to intensity based registration has the following concerns to be addressed:

i. Supervised or unsupervised, ideally unsupervised would be preferred to avoid any kind of manual intervention in the registration process.
ii. The accuracy and the speed of the segmentation, it needs to be established that how much accuracy in segmentation is mandatory for robust performances.

iii. Segmentation required only in one image or both the images involved in the registration process.

The proposed method is unsupervised, very fast and easy to implement and requires segmentation in only one of the images being registered (reference image preferably to reduce computation time). Inspired from the results obtained for dataset 1, the idea of the proposed solution lies in the histogram of a SAR image acquired over urban/semi-urban areas. Normally, the pixels generated by the double/triple bounce phenomenon result into a very strong backscatter to the radar sensor and thus most of these pixels would be located towards the higher end of a SAR image histogram.

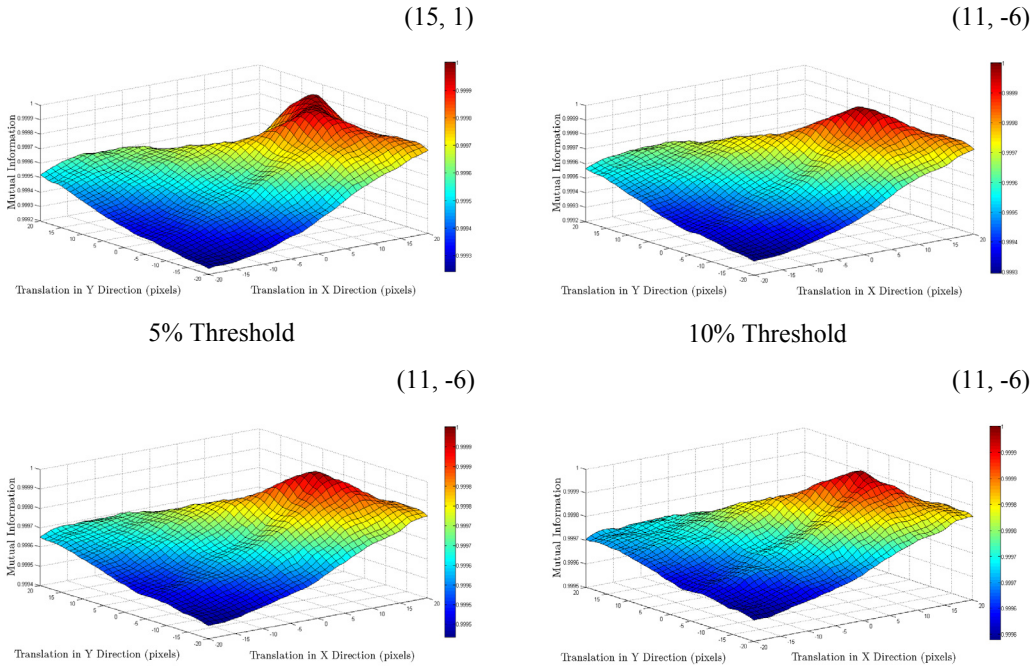


Figure 4.10: Registration surfaces generated by MI between segmented SAR (using masks in Figure 4.9) and the original optical image. As threshold level reaches 20-30% we obtain the similar peak as was generated by plain field pixels (Figure 4.8). The x and y axis represent translation values of [-20, +20] pixels in x and y direction respectively. The z axis represents the normalized MI values obtained. The peak of the registration surfaces are mentioned on the top right

Here it has to be kept in mind that certain other pixels (not influenced by SAR geometry) due to constructive interference of the radar waves can also produce high intensity value (strong backscatter). However by histogram thresholds, it still might be possible to bin out most of the

pixels explicitly produced by the SAR sensor working principle. For dataset 1, MI recorded the same registration peak for different threshold levels so as long as the numbers of "false" pixels being binned out represent minority of the total pixel population, the true registration peak is not expected to change. The number of such false pixels can definitely be reduced by speckle filtering but intensity based registration of SAR and optical imagery does not require any mandatory image smoothing step, hence we refrain to perform the same in this solution.

The results of the proposed segmentation scheme can be visualized in Figure 4.10. First the SAR image (Figure 4.8b) is down-sampled to one-forth of its original resolution. The histogram of the obtained down-sampled image is now used to generate thresholds for binning out possible pixels affected by the SAR sensor geometry in the original resolution image. To realize the goal of the segmentation process, thresholds are made from the higher end of the image histogram. It is clearly observed in Figure 4.9 that as the threshold limit is relaxed (from 5 to 30%) more and more pixels from the plain fields start coming into the filtered pixel category and this might have an adverse influence on similarity metric performance.

To analyze the segmentation influence, we register the segmented SAR image (different thresholds analysed) with the corresponding optical image. In this scenario, all the pixels from optical imagery would contribute to the similarity metric computation but from the SAR imagery only those pixels which are within the threshold limits (assigned the value 0 in the masks of Figure 4.9) would participate in the registration process. The MI surfaces generated in the search space of [-20 20] pixels for the segmented SAR images and the original optical image have been provided in Figure 4.10. Segmentation of the SAR image using the mask depicted in Figure 4.9a (5% threshold) influenced the registration peak observed in Figure 4.8c. The registration peaks obtained by MI (17, 4) shifts to (15, 1). Further segmentation of the SAR image i.e. using threshold in the order of 10, 20 and 30% percent yielded nearly the same registration peaks as were reported by the similarity metrics using only from plain field pixels (Figure 4.8d). The MI peaks obtained for the segmented SAR and the optical imagery are same as the peaks obtained earlier (1 m difference in y direction still exist) using only the plain field pixels, which in this case are assumed to be the true registration parameters.

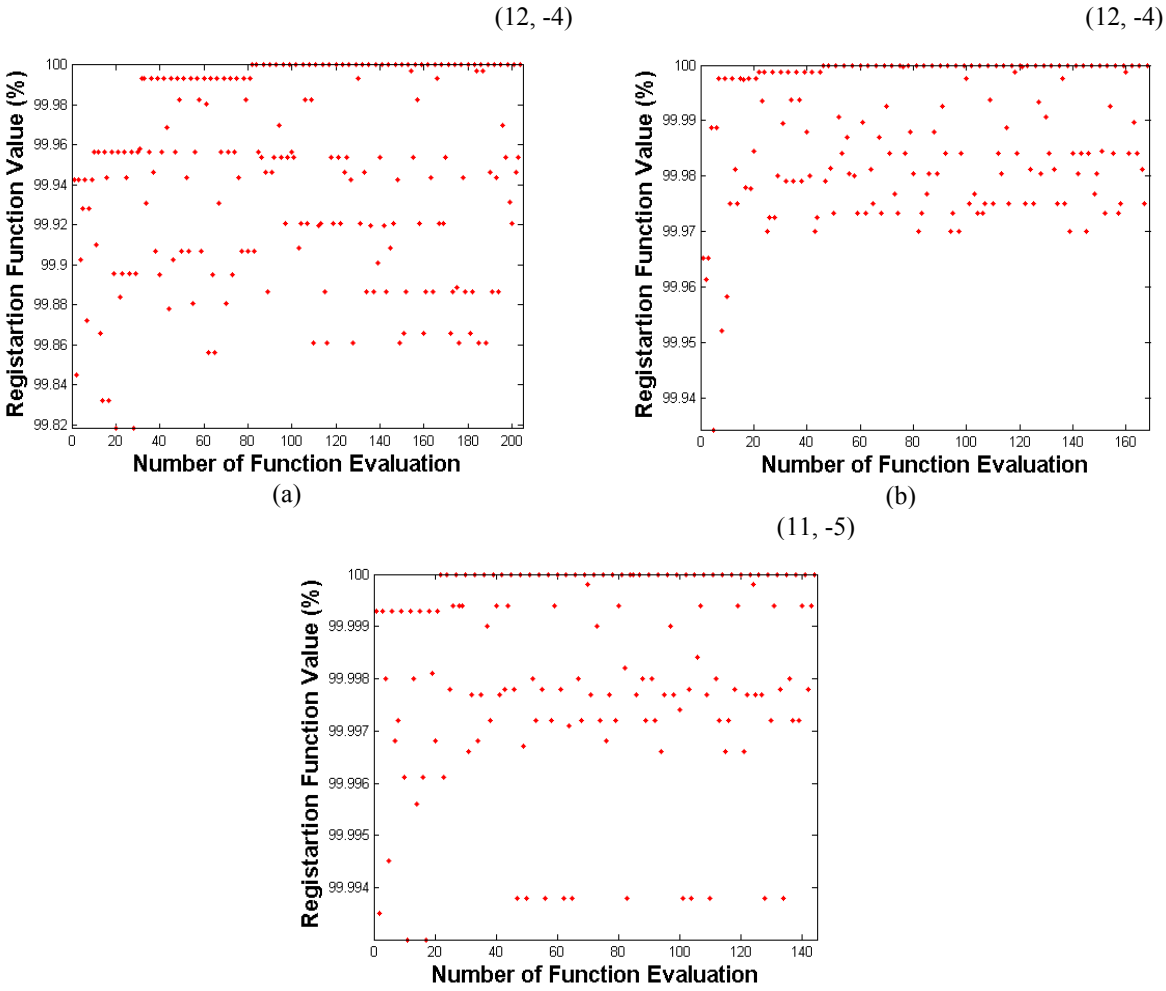


Figure 4.11: Performance of the discrete SPSA algorithms for segmented TerraSAR-X and IKONOS imagery, (a-c) represent MI function optimization at 4 m 2 m and 1 m pixel spacing respectively using nearest neighbour technique (obtained registration parameters reported on top right of the optimization graphs).

Further, the performance of the discrete SPSA algorithm can be visualized in Figure 4.11. Analyzing the search spaces in Figure 4.10 the registration peak can be expected to be around (11, -6) pixels in x and y direction for 20% thresholded TerraSAR-X and the IKONOS image. The seed utilized for the optimizer is (0, 0) and we perform out optimization in multiresolution pyramid comprising of images having 4 m, 2 m and 1 m pixel spacing. Figure 4.11a-c depict performance of the discrete SPSA using nearest neighbour technique from 4 m to 1 m images respectively. Here at 1 m pixel spacing, we obtain (11, -5) pixel translation in x and y direction as the registration parameters which differ by 1 m in y direction from the expected result.

4.6 CONCLUSION

In this chapter we have analysed the performance of mutual information for metric resolution images acquired by SAR and optical sensors. The two major concerns with high resolution data are incompatible imagery content and high data volume. The incompatible radiometric information and geometric behaviour within the images do influence the mutual information performance but the introduced intensity based thresholding in the SAR image has been found beneficial to improve the similarity metric performance. Further examples shown in the subsequent chapter show that this method gives encouraging results for various tested datasets. The other issue of high data volume generally encountered in remote sensing image registration can be handled by using the proposed discrete optimization scheme to reduce the registration processing time considerably. In the following chapter, we present a critical analysis of mutual information for achieving registration of high resolution imagery acquired over dense urban areas.

CHAPTER 5

MUTUAL INFORMATION BASED REGISTRATION OF TERRASAR-X AND IKONOS IMAGERY IN URBAN AREAS

5.1 INTRODUCTION

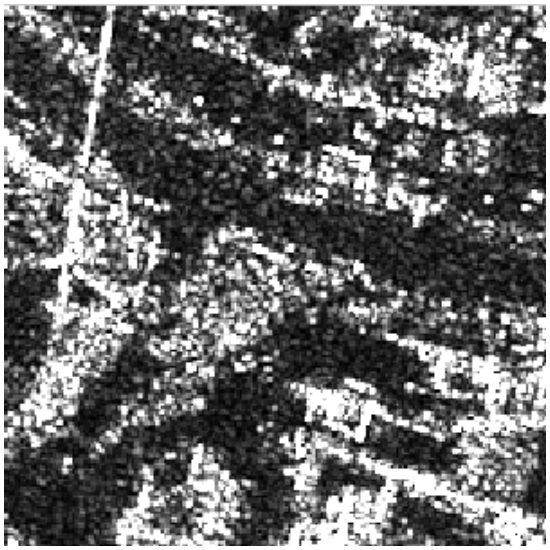
The previous chapter highlighted a possible solution to the problematic behaviour of mutual information while utilizing high resolution imagery especially acquired over scenes having a significant number of real world 3D objects. In this chapter, we test dense urban area images where different sensor characteristics have a major influence on the performance of the investigated similarity metric. We analyse techniques, issues and possible solutions to adapt mutual information for registration of high resolution imagery acquired over dense urban areas. Further, based on the analysis in this and the previous chapter, we present a novel method to improve the sensor orientation for orthorectification of high resolution optical data.

5.2 CHALLENGES IN REGISTERING HIGH RESOLUTION URBAN AREA IMAGES

Automatic registration of high resolution data is a tough task due to two reasons, one is data volume and other being the amount of fine details now visible in satellite imagery. For multisensor images the situation further complicates because of possibly different modality and/or geometry of acquisition (see Figure 5.1). Multisensor modality can be handled by registration similarity metrics but for scenes (like those acquired over dense urban areas) where different sensor geometries and imaging principles start playing a crucial role, pose much more challenges to all registration techniques. In Figure 5.1, we have three images one from TerraSAR-X (acquired 2 days after the earthquake on 12/05/2008 in Sichuan, China) and two IKONOS images acquired pre and post the disastrous natural calamity.

The challenges associated in registering each of the IKONOS images to the TerraSAR-X image can simply be visualized by the very different appearance of the images acquired by the

sensors. Not only the SAR image offers completely different signatures of the 3D urban settlements, the 2D features like the main road which is clearly observed in the optical images is difficult to interpret and delineate in case of the SAR image.



(a)

(b) (c)

Figure 5.1: Challenges in registering high resolution images acquired over urban areas in form of different sensor radiometry and geometry. (a) TerraSAR-X image acquired after 2 days of the earthquake. (b) IKONOS image acquired 8 months before the earthquake. (c) IKONOS image acquired a month after the disaster. Both IKONOS images have different geometry of acquisition. Red cross indicates a "on ground" feature point detected and matched automatically within the IKONOS images.

Looking at the SAR-optical image pairs it is a subtle observation that developing a feature point algorithm to detect and match similar features might just not be possible. The most realizable solution is to use intensity based techniques and try to increase the influence of on ground features in the matching process. The accuracy of this method with the proposed attempt is analysed later in this chapter.

Table 5.1: Details of TerraSAR-X and IKONOS imagery utilized for analysis in dense urban areas.

	TerraSAR-X	**IKONOS-2** (pre disaster)	**IKONOS-2** (post disaster)
Mode	High resolution spot light (HS) Descending orbit	Forward scanning	Forward scanning
Spectral Resolution	9.65 GH	450 - 900 nm	450 - 900 nm
Spatial Resolution	Range: 1.64 m Azimuth: 1.66m	Cross Scan: 1.03 m Along Scan: 0.92 m	Cross Scan: 0.94 m Along Scan: 1.08 m
Geometry	Incidence angle: 46.8º	Sun azimuth: 154.4 º Sun elevation: 60.3º Collection azimuth: 284.4º Collection elevation: 62.1º	Sun azimuth:113.0º Sun elevation:73.3º Collection azimuth: 352.8º Collection elevation: 59.3º
Bits per pixel	16 bit	11 bit	11 bit
Date of Acquisition	15/05/08	14/09/07	28/06/08
Product	GEC Product	IKONOS GEO	IKONOS GEO

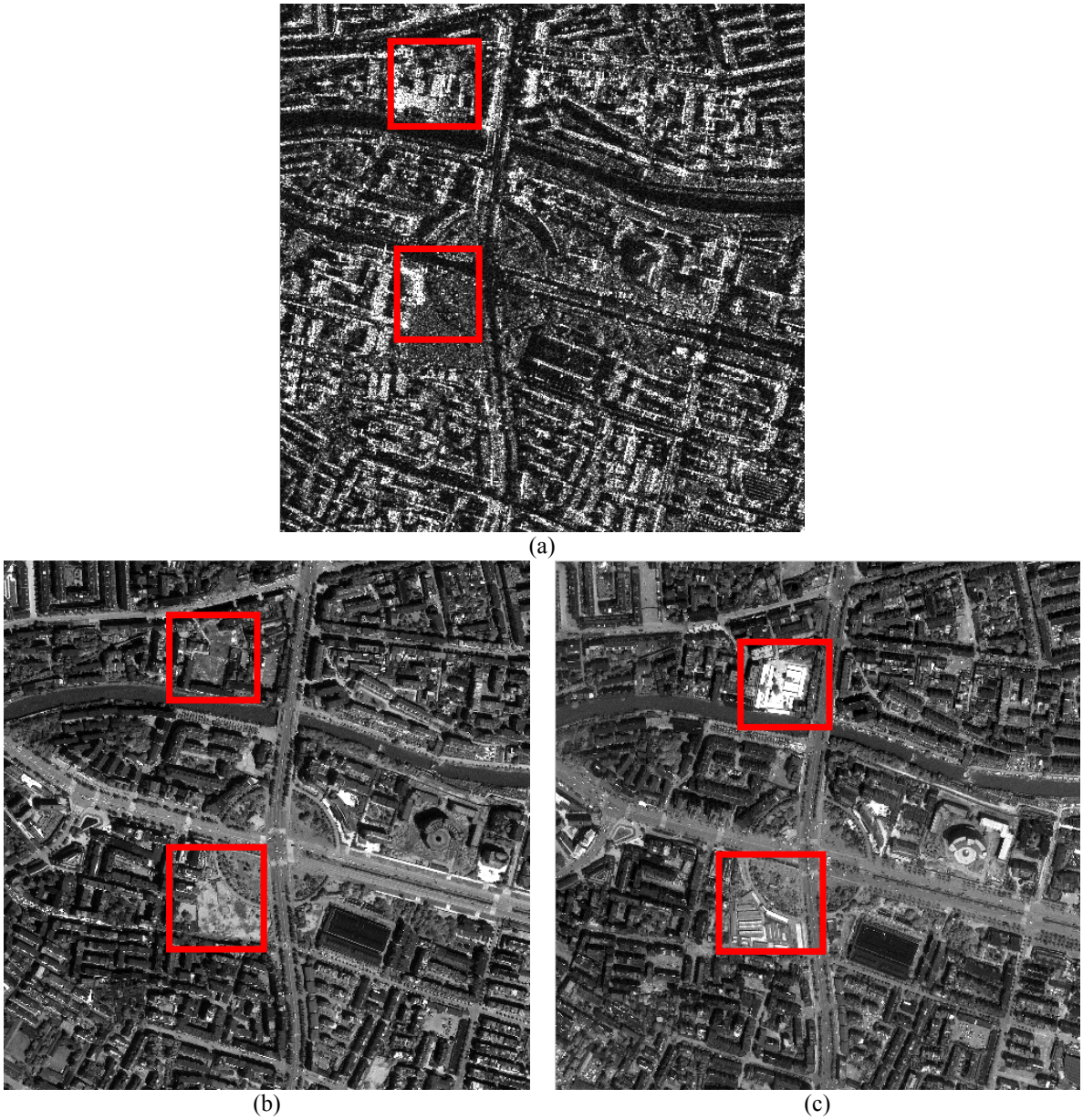
Considering the two IKONOS images shown in Figure 5.1, it becomes clear that images even for the same sensor but acquired under different geometrical conditions produce different signatures for 3D objects. This is observed by analysing the shadow responses of the different buildings shown in the optical images. Specifically for these IKONOS images, different shadow pattern is a result of different geometrical settings (collection azimuth and collection elevation) and sun position (sun azimuth and sun elevation) during the time of image acquisition. To understand how these mentioned angles influence the geometry of acquisition for IKONOS images, interested readers are referred to Grodecki and Dial (2001). Further, as these images have similar modalities we can use point feature detectors and matching schemes (like SIFT) but specifically for urban areas, after the matching phase, we would require a classification of the matched features into “on ground” and “above ground” features as for registration purposes only the former can be utilized. Generally, urban features like buildings, object shadows, small real world occlusions would make this kind of classification difficult to implement. A sample “on ground” feature (relevant for image registration) detected by the SIFT operator for these image

chips is marked by a red cross in the two images (Figure 5.1). Although, high resolution optical image registration is not the theme of focus in this dissertation, it is anticipated that very high resolution optical urban area scenes even acquired by the same sensor but under different geometrical settings might pose some challenges for similarity metrics like mutual information.

5.3 REGISTRATION IN URBAN AREAS

In this section, we present a scenario that the end users might be confronted with while utilizing high resolution images acquired over urban areas. In the context of the presented work, registration is defined as estimating shifts in x and y direction between already orthorectified imagery. This also implies that the discrete optimization scheme proposed in the previous chapter can be very well utilized here to quickly estimate large global shifts between voluminous high resolution images. We consider TerraSAR-X and IKONOS imagery acquired over the city of Sichuan in China (dataset details in Table 5.1). The pre disaster image has been acquired roughly eight months prior to the earthquake and the post disaster IKONOS image is acquired approximately a month after. The image pairs have been procured from georeferenced scenes (without DEM) and have considerable urban changes within the acquired time period. The TerraSAR-X image is the standard GEC product which has been corrected to UTM projection using a constant ellipsoidal height and is not the most accurate data product available (Fritz and Eineder, 2008). Moreover, the following georeferencing differences within the images are observed (based on average of few manual measurements):

i. TerraSAR-X and the post disaster IKONOS image have a georeferencing difference of approximately 90 m in x direction and 45 m in y direction.
ii. TerraSAR-X and the pre disaster IKONOS image have a georeferencing difference of approximately 125 m in x direction and 25 m in y direction.
iii. The average shift within the two IKONOS image using the SIFT matching technique is estimated as 37 m in x direction and 20m in y direction with 1 m and 2 m standard deviation in x and y directions respectively (from 8 automatically matched “on ground” points), reported parameters have been rounded to nearest integers and are used in experiments later for evaluating registration consistency



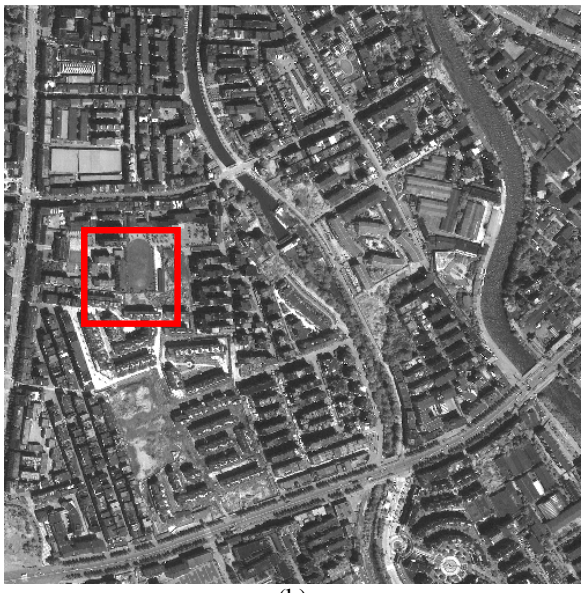
(a)

(b) (c)

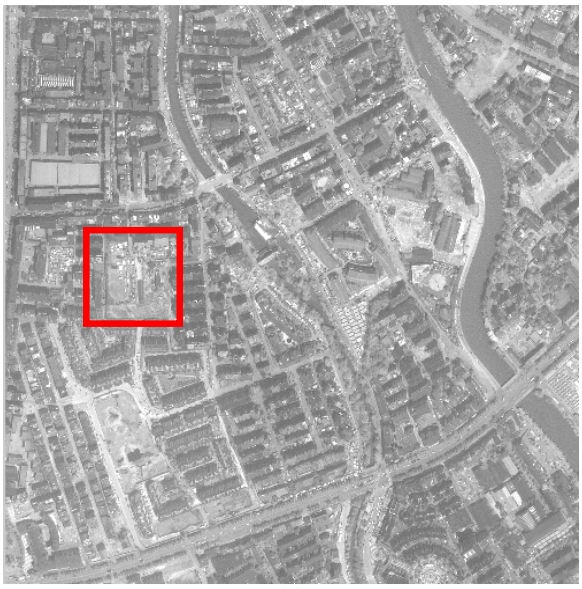
Figure 5.2: Images utilized in dataset-1a. (a) TerraSAR-X image acquired after 2 days of the earthquake. (b) IKONOS image acquired 8 months before the earthquake. (c) IKONOS image acquired a month after the disaster. Markings reflect the areas in the images that have gone under considerable urban change within the acquisition time span.

For complete analysis, we retrieve shift parameters between two datasets (zero overlap) extracted from the same complete scene (image pairs of size 1000x1000 pixels at 1 m pixel spacing are referred to as dataset-1a and 1b in further text and are provided in Figure 5.2 and 5.3 respectively). Here the shifts are estimated keeping the TerraSAR-X image as the reference and the two IKONOS images as the input image.

(a)



(b)



(c)

Figure 5.3: Images utilized in dataset-1b. (a) TerraSAR-X image acquired after 2 days of the earthquake. (b) IKONOS image acquired 8 months before the earthquake. (c) IKONOS image acquired a month after the disaster. Markings reflect the areas in the images that have gone under considerable urban change within the acquisition time span.

The registration involving the pre and post disaster IKONOS image are referred as "Pre Registration Scenario" and "Post Registration Scenario" in further text.

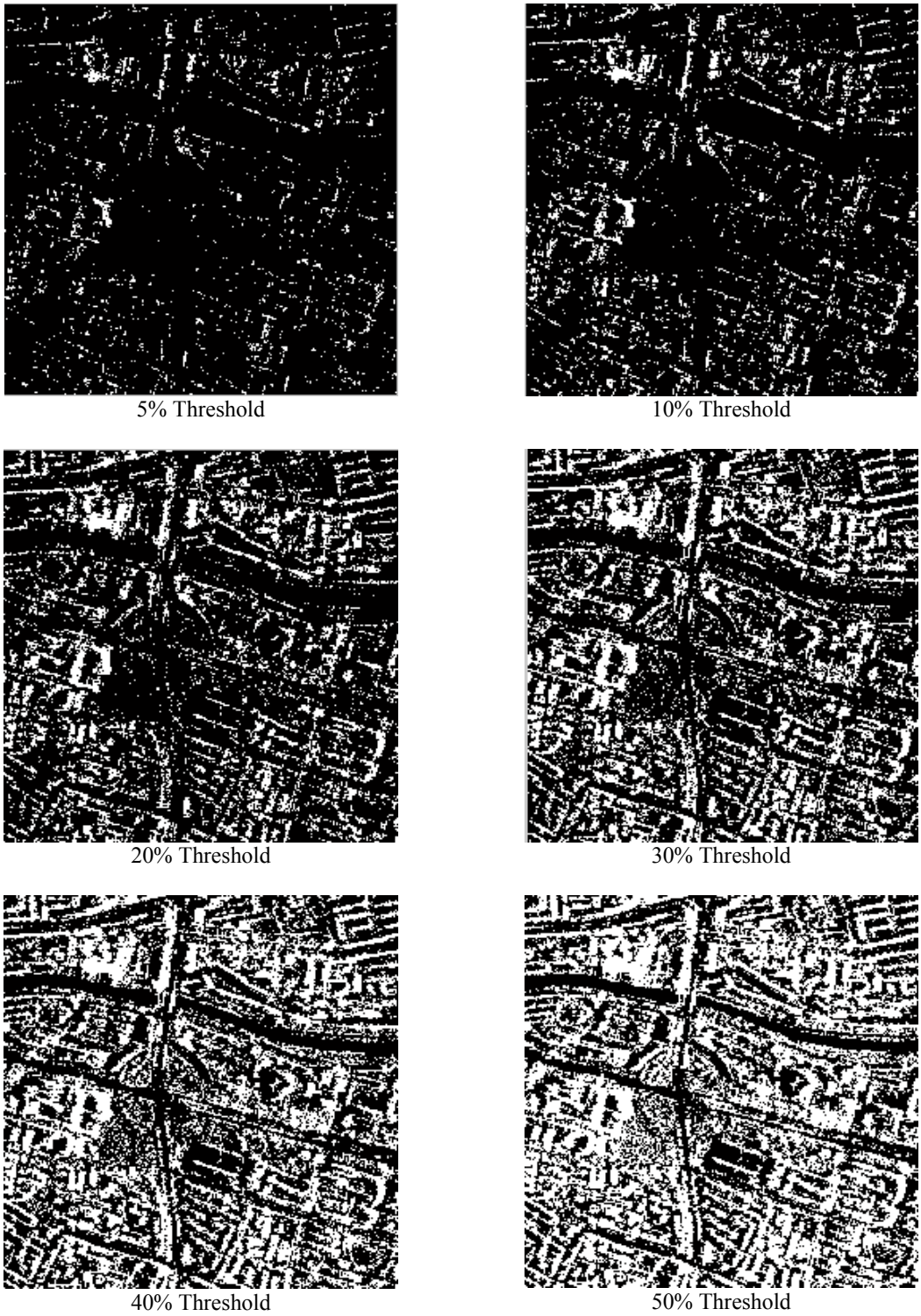


Figure 5.4: Thresholded TerraSAR-X image of dataset-1a. Pixels with value 1 are left out (in SAR image) of the registration process after introducing high thresholds of 5 to 50% at image compressed to one-forth of its original resolution. It is observed that most of the "on ground" pixels are left for matching after the proposed thresholding step.

It is worthwhile to mention that the above mentioned georeferencing differences are not artificially introduced but are originally present in the procured scenes. The acquisition time differences within the images and the aftermath of the disastrous earthquake have led to significant urban change in the analysed datasets. We observe new roads, buildings and open spaces especially full of temporary settlements in the post earthquake IKONOS image. Some of the profound changes within the three images are demarcated in Figure 5.2 and 5.3. In this chapter, we analyse two registration techniques, one technique is based on the thresholding scheme proposed in the previous chapter and the other technique is based upon a region growing methodology (Türmer, 2009) that shows much potential for high resolution image matching/registration applications.

5.3.1 Histogram Based Registration

To continue evaluating the thresholding scheme proposed in the previous chapter, we translate the IKONOS image over the reference TerraSAR-X image in a range of -120 to -80 meters in x and -65 to -25 meters in y direction (post disaster) and -150 to -110 meters in x and -50 to -10 meters in y direction (pre disaster). The true registration parameters for both the image pairs are expected to lie within this range and have been estimated manually by marking conjugate control points. We perform this experiment within images having pixel spacing of four, two and 1 m respectively. Objective here is to do a multiresolution consistency analysis of the mutual information metric with the proposed thresholding scheme. Unless and otherwise stated, in this section all the computations have been made from a 128 bin size joint histogram estimated through cubic B-spline kernel and the image pyramid is built using the mean block filter.

Results from dataset-1a

To analyse the influence of the SAR image pixels generated mainly by double bounce effect in urban areas we performed the same segmentation scheme tested effectively for sub urban images in the previous chapter. The segmented TerraSAR-X image with thresholds in range 5 to 50% can be visualized in Figure 5.4. We first test the metric performance for the original images and later for segmented SAR and the optical image at pixel spacing of four, two and one meter.

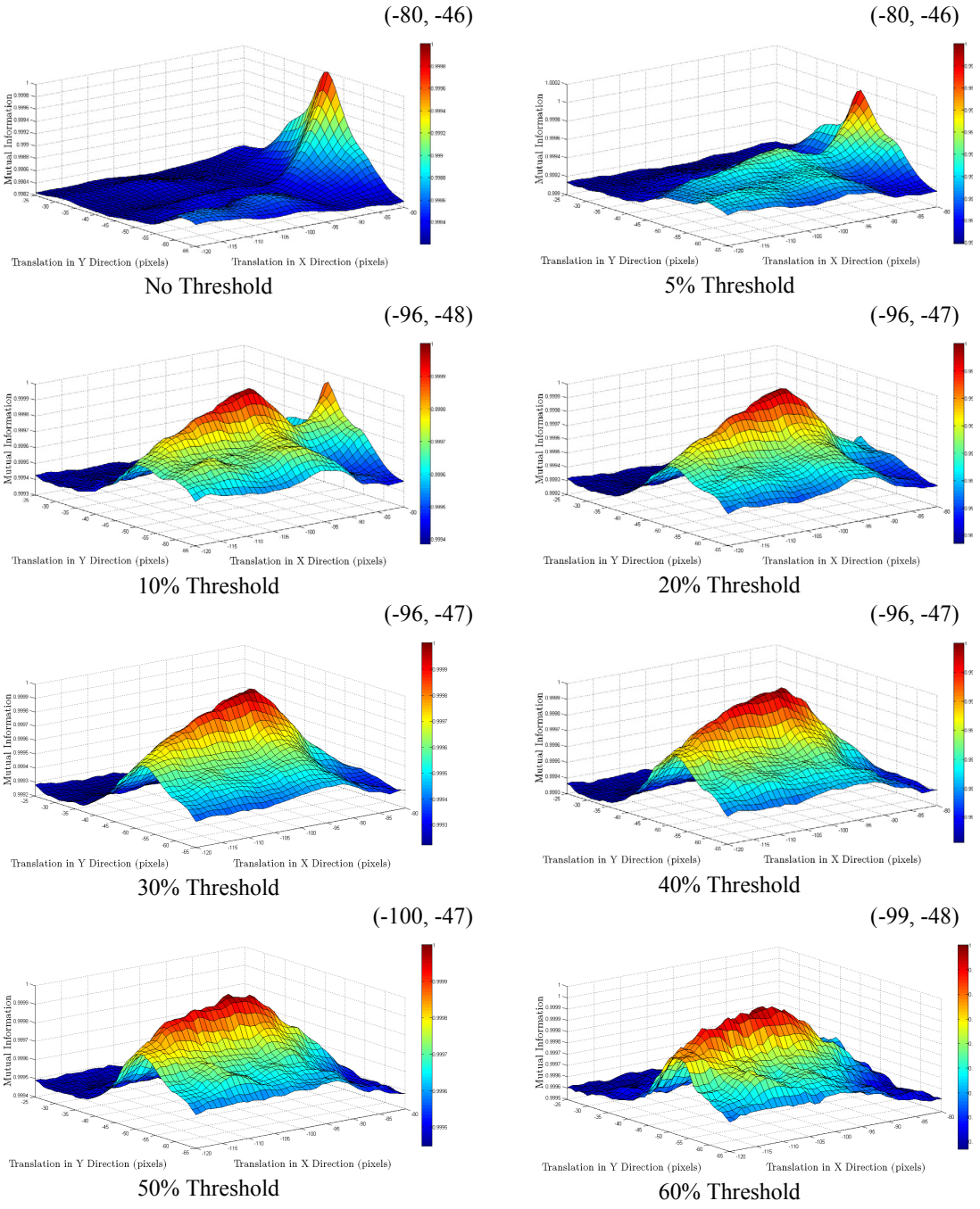
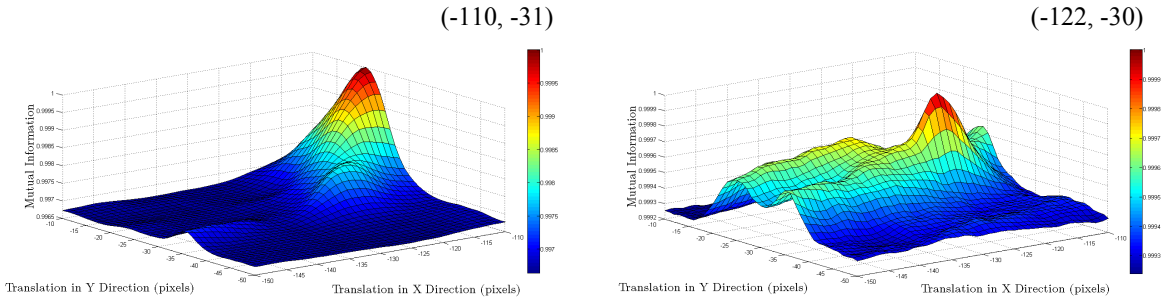


Figure 5.5: Registration surfaces generated by MI (dataset-1a, post disaster scenario) between segmented SAR and the original optical image. Interpolation artefacts may increase with the level of threshold in SAR image (Suri and Reinartz, 2009). The x and y axis represent translation values of [-120, -80] pixels in x and [-65, -25] pixels in y direction respectively. The z axis represents the normalized MI values obtained. The peak of the registration surfaces are mentioned on the top right



(a) (b)

Figure 5.6: Registration surfaces generated by MI (dataset-1a, pre disaster scenario) between (a) original SAR and the original optical image (b) 30% thresholded SAR and optical image. The x and y axis represent translation values of [-150, -110] pixels in x and [-50, -10] pixels in y direction respectively. The z axis represents the normalized MI values obtained. The peak of the registration surfaces are mentioned on the top right

Table 5.2: Multiresolution performance of MI for registration of pre and post disaster IKONOS imagery to TerraSAR-X image of dataset-1a (Figure 5.2).

		L2	**L1**	**L0**
Pixel Spacing		4 m	2 m	1 m
Pre Search Space (pixels)		x: -37 to -22 y: -12 to -2	x: -75 to -55 y: -25 to -5	x: -150 to -110 y: -50 to -10
Post Search Space (pixels)		x: -30 to -20 y: -17 to -7	x: -60 to -40 y: -33 to -13	x: -120 to -80 y: -65 to -25
Registration parameters in meters (x and y direction)				
Original Image Pairs (100% entropy)	Pre	-112, -22	-110, -21	-110, -31
	Post	-80, -46	-80.0, -46	-80, -46
5% High Threshold (97% entropy)	Pre	-112, -21	-110.0, -22	-110, -31
	Post	-80, -46	-96.0, -48	-80, -46
10% High Threshold (95% entropy)	Pre	-112, -22	-110.0, -22	-110, -31
	Post	-96, -48	-98.0, -48	-96, -47
20% High Threshold (92% entropy)	Pre	-120, -28	-122.0, -30	-122, -30
	Post	-96, -48	-98.0, -48	-96, -47
30% High Threshold (89% entropy)	Pre	-132, -28	-122.0, -30	-122, -30
	Post	-96, -47	-98.0, -48	-96, -47
40% High Threshold (86% entropy)	Pre	-132, -28	-122, -30	-122, -30
	Post	-96, -47	-100, -48	-96, -47
50% High Threshold (84% entropy)	Pre	-132, -28	-124, -30	-122, -30
	Post	-100, -47	-100, -48.	-100, -47

The peaks observed for the two image pairs along with the TerraSAR-X image entropy content have been tabulated in Table 5.2. The same entropy behaviour as highlighted for dataset-1 of the previous chapter is depicted by this TerraSAR-X urban area image with approximately high 50% intensity values only constitute 16% of the entire image information. As observed earlier, here also the registration peak obtained for the original imagery shifts as the pixels influenced by the

SAR image acquisition principle are removed from similarity metric computation. For the 1 m pixel spacing post disaster IKONOS and the TerraSAR-X image pair (dataset-1a, Figure 5.2), MI reports registration peak at (-80, -46) for original images and this peak is shifted to (-96, -47) on introducing thresholds of 10 to 40% in the SAR image. A similar trend is observed for the 2 m and 4 m pixel spacing images. For the pre-disaster case, MI reports a registration peak at (-110, -31) pixels for the original image pairs and finally registration peaks stabilize at (-122, -30) pixels for threshold ranging from 20 to 50% in the SAR image. A similar trend reflecting shifts in MI peaks is also observed at 2 m and 4 m pixel spacing images. MI behaviour for three resolution levels using different thresholds is tabulated in Table 5.2.

Specifically for the post disaster IKONOS and the SAR image, an observation of the generated MI search spaces in Figure 5.5 clearly indicates a systematic shift of registration peak from (-80, -46) to (-96 -47) pixels in x and y direction as the image thresholds are increased from 5 to 40%. For the original images and 5% high thresholded SAR image the peak is at (-80, -47) pixels. A small local maximum is observed around (-96, -47) is observed for 5% threshold image which becomes more profound for 10% thresholded SAR image. Further thresholding of the SAR image (till 40%) makes the registration peak stable at (-96, -47) pixels in x and y direction. As further information in form of pixels taking part in the registration process from the SAR image are removed (threshold 50 and 60%) the MI registration search space is observed to be rough and the peak also show small shifts in both x and y directions. The search space for the pre disaster scenario with similar trend is depicted in Figure 5.6.

To do a consistency analysis of the MI performance, we manually estimate the rough registration shifts between the registered pre and post disaster IKONOS images by parameters obtained at different pixel spacing. With an assumption that after thresholding MI metric provides better performance, we estimate registration errors on basis of manual observations within the registered images. The individual *Deviation* estimate is reported using the following relations

$$Deviation_x = Manual_x - MI_x$$

$$Deviation_y = Manual_y - MI_y$$

$$Deviation = \sqrt{Deviation_x^{\,2} + Deviation_y^{\,2}} \tag{5.1}$$

Where $Manual_x$ and MI_x represent the parameters estimated manually and using MI respectively in the x direction, similar notation is also valid for the y direction. *Deviation* is the absolute difference between the mutual information estimate and the manually measured registration shifts.

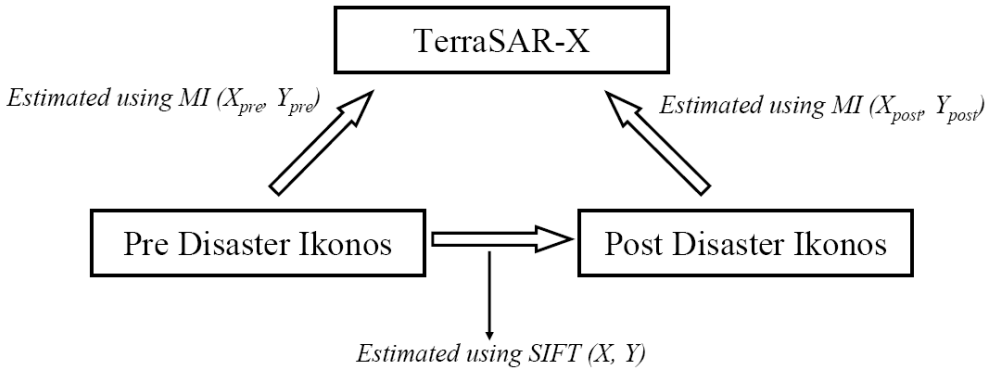


Figure 5.7: Evaluation of the mutual information performance consistency.

As we already have estimates of the shifts present within the two IKONOS images (see Figure 5.7), Registration Consistency RC is also utilized to evaluate MI performance and is estimated using the following equations

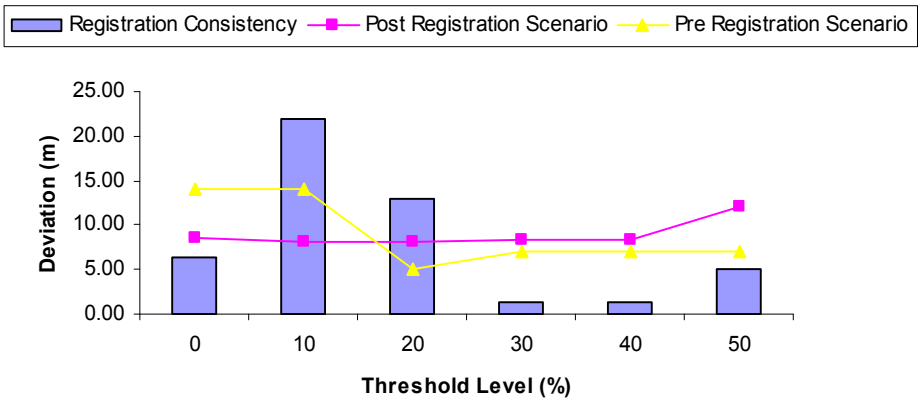
$$RC_x = X_{SIFT} - X_{pre} + X_{post}$$

$$RC_y = Y_{SIFT} - Y_{pre} + Y_{post}$$

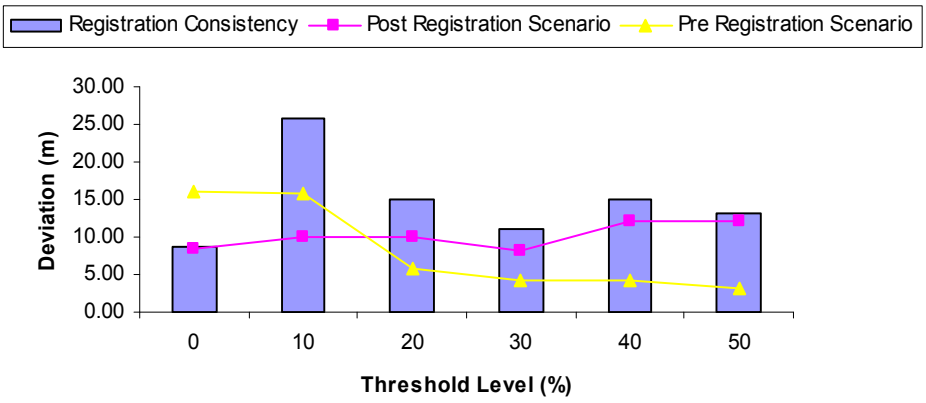
$$RC = \sqrt{RC_x^{\,2} + RC_y^{\,2}} \tag{5.2}$$

In the above equations $X_{pre}, X_{post}, X_{SIFT}$ represent the x direction registration estimate for the pre disaster IKONOS image to TerraSAR-X (estimated using MI), post disaster IKONOS image to TerraSAR-X (estimated using MI), and pre disaster IKONOS to post disaster IKONOS image (estimated using SIFT). Similar notation is used for the y direction as well. Finally we compute the absolute RC measure using its x and y component.

(a) 4 m pixel spacing



(b) 2 m pixel spacing



(c) 1 m pixel spacing

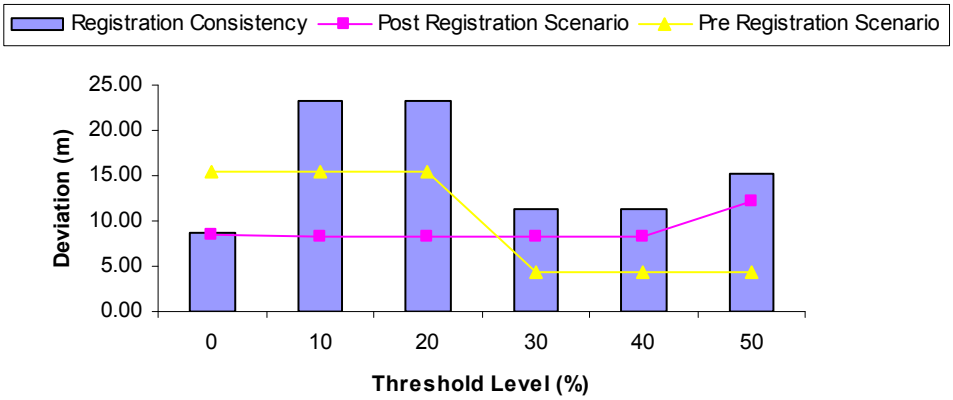


Figure 5.8: Mutual information performance using the thresholding scheme at different pixel spacing for dataset-1a (Figure 5.2). Registration consistency is computed using Equation 5.2 and individual deviation for post and pre registration scenarios are computed using Equation 5.1.

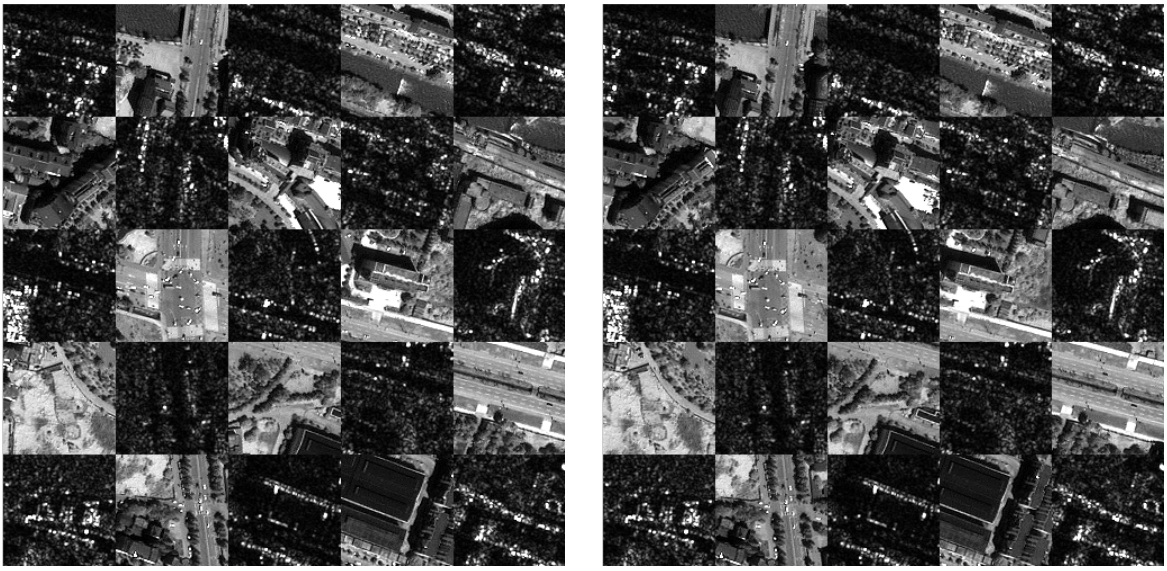


Figure 5.9: TerraSAR-X image (Figure 5.2) check-squared with registered pre disaster IKONOS image. We have utilized before thresholding (-110, -30) and after thresholding (-122, -30) parameters for the left and the right images respectively. The difference in registration is very hard to observe and is subjective to observers experience in absence of ground truth. The influence of thresholding in terms of improved registration is much clearer in results of Figure 5.12.

Ideal value depicting perfect MI performance would be zero for the RC measure but it does not necessarily ensure the correct registration parameters for individual registration scenarios. Therefore, individual registration errors estimated considering the manual observations need to be incorporated for an objective analysis. For this particular case achieving a RC value of zero might be difficult as the two IKONOS images involved have been acquired with different geometrical configurations (sun azimuth and elevation, collection azimuth and elevation) and this might influence certain features especially roads within building lanes. Ideally for this kind of analysis one needs two IKONOS scenes acquired at different dates but with same geometrical settings.

The registration parameters estimated manually for pre and post registration scenario for dataset-1a are (-125, -27) and (-88, -49) respectively (rounded to nearest integers) these have been obtained by manually marking four control points at a standard deviation of 2-3 m in x and y directions. The individual registration consistencies RC and expected accuracy of the pre and post disaster registration results tabulated in Table 5.2 can be visualized in Figure 5.8(a-c) for pixel spacing of four, two and one meter respectively. Results in Figure 5.8 suggest that the post

disaster registration scenario has a deviation of around 8 m where as on the other hand MI reports 5 m deviation for the pre disaster registration scenario at three different resolution levels. It is also observed that the introduced threshold did not improve the deviation for the post disaster scenario (remains around 8 m although the MI peak shifts) but has a significant influence in the pre disaster registration scenario (deviation reduced from 15 m to 4 m). The best results for this dataset seem to be achieved at 30% threshold at all the three resolution levels as we observe the minimum RC measure value as well as minimal deviation for the two scenarios. The registered TerraSAR-X and pre disaster IKONOS imagery utilizing parameters before and after the thresholding process can be visualized in Figure 5.9. As for this dataset, there has not been a huge shift in registration peaks, visualizing the differencing within the two images is a tough ask. To further investigate the mutual information performance and the thresholding scheme, we estimate registration parameters in dataset-1b which has images from the same complete scene.

Results from dataset-1b

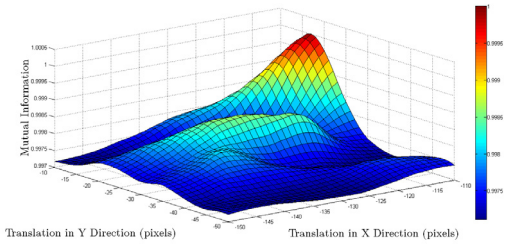
For dataset-1b, the selected sub scenes are provided in Figure 5.3. In this case the MI metric gives similar trends as observed for dataset-1a to yield the observation table in Table 5.3. For this case the post disaster registration scenario shows a peak shift from (-80, -39) to (-94, -50) after thresholding and on similar lines the pre disaster registration scenario shows a peak shift from (-110,-17) to (-123, -29) (till 40%) at 1 m pixel spacing. Thresholding shifts the MI peak in both x and y direction as compared to the previous case where it dominantly only influenced the x direction. The MI search spaces both before and after the thresholding at different resolution levels are provided in Figure 5.10. It is seen that at 4 m pixel spacing, on plotting the MI surface interpolation artefacts appear for pre and post disaster registration scenarios. These can be attributed to the introduced thresholding and the performed sub pixel shifts. However, artefacts are removed by using the Sextic B-spline kernel as shown in Figure 5.10. This example further validates the effectiveness of the Sextic B-spline kernel for handling interpolation artefacts in high resolution image registration applications (Suri and Reinartz, 2008).

Table 5.3: Multiresolution performance of MI for registration of pre and post disaster IKONOS imagery to TerraSAR-X image of dataset-1b (Figure 5.3).

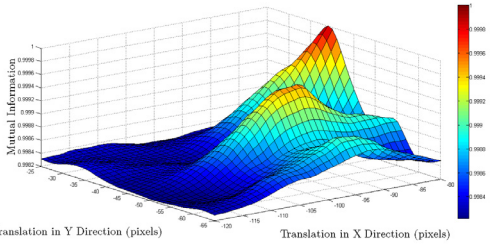
		L2	**L1**	**L0**
Pixel Spacing		4 m	2 m	1 m
Pre Search Space (pixels)		x: -37 to -22 y: -12 to -2	x: -75 to -55 y: -25 to -5	x: -150 to -110 y: -50 to -10
Post Search Space (pixels)		x: -30 to -20 y: -17 to -7	x: -60 to -40 Y: -33 to -13	x: -120 to -80 y: -65 to -25
Registration parameters in meters (x and y direction)				
Original Image Pairs (100% entropy)	Pre	-131, -27	-110, -17	-110, -17
	Post	-95, -50	-94, -50	-80, -39
5% High Threshold (97% entropy)	Pre	-128, -28	-132, -26	-123, -29
	Post	-92, -52	-94, -50	-94, -50
10% High Threshold (96% entropy)	Pre	-128, -28	-130, -26	-124, -29
	Post	-92, -52	-94, -50	-94, -50
20% High Threshold (92% entropy)	Pre	-128, -28	-130, -26	-124, -27
	Post	-92, -52	-94, -50	-94, -50
30% High Threshold (90% entropy)	Pre	-128, -28	-132, -26	-123, -29
	Post	-92, -52	-94, -50	-94, -50
40% High Threshold (87% entropy)	Pre	-136, -24	-132, -26	-123, -29
	Post	-96, -52	-96, -50	-94, -50
50% High Threshold (84% entropy)	Pre	-136, -24	-132, -26	-132, -26
	Post	-96, 52	-96, -50	-94, -50

The registration parameters estimated manually for pre and post registration scenario for dataset-1b are (-125, -30) and (-89, -48) respectively (estimated using 4 control points with approximately 3 m standard deviation in both x and y direction). Assuming these to be correct parameters, we have deviation of around 6 m and 2 m for the post disaster and pre disaster registration scenario (Figure 5.11), which suggests a much better performance than the previous case. The TerraSAR-X image registered with the pre disaster IKONOS images using the parameters obtained before and after the thresholding scheme are presented in Figure 5.12.

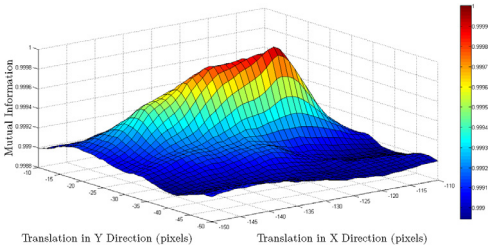
No Threshold at 1 m (-110, -17)



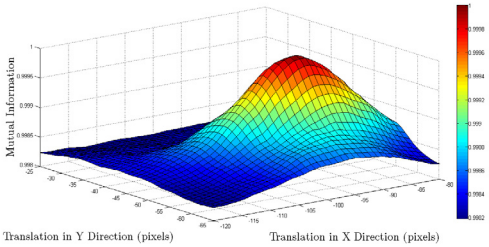
No Threshold at 1 m (-80, -39)



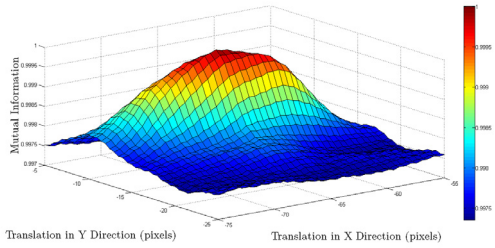
30% Threshold at 1 m (-123, -29)



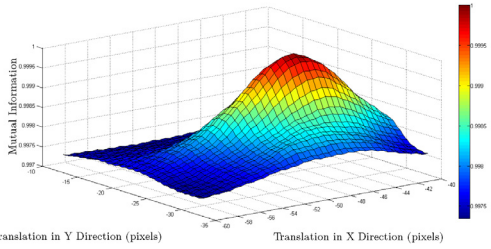
30% Threshold at 1 m (-94, -50)



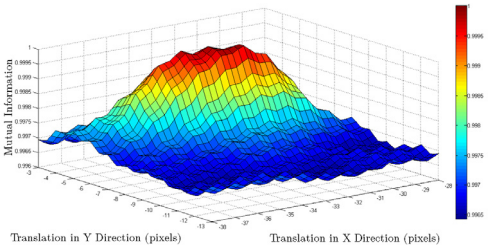
30% Threshold at 2 m (-132, -26)



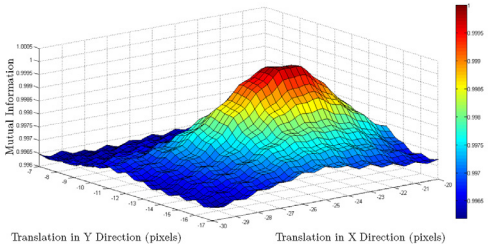
30% Threshold at 2 m (-94, -50)



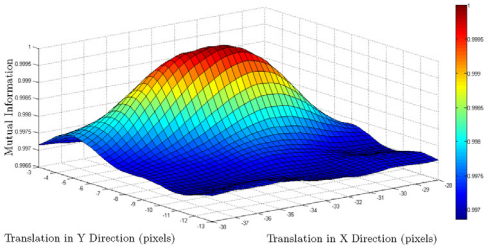
30% Threshold at 4 m (-128, -28)
Cubic B-spline



30% Threshold at 4 m (-92, -52)
Cubic B-spline



30% Threshold at 4 m (-128, -28)
Sextic B-spline



30% Threshold at 4 m (-94, -50)
Sextic B-spline

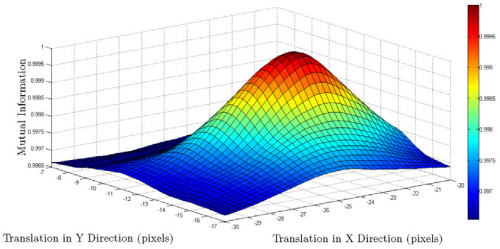
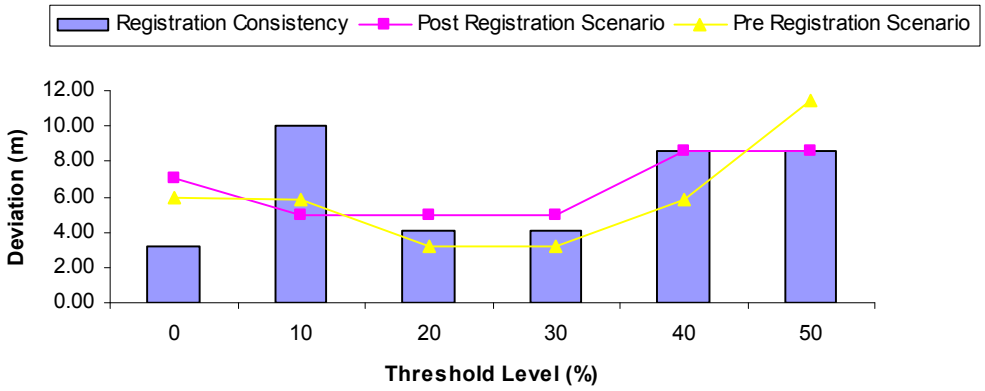
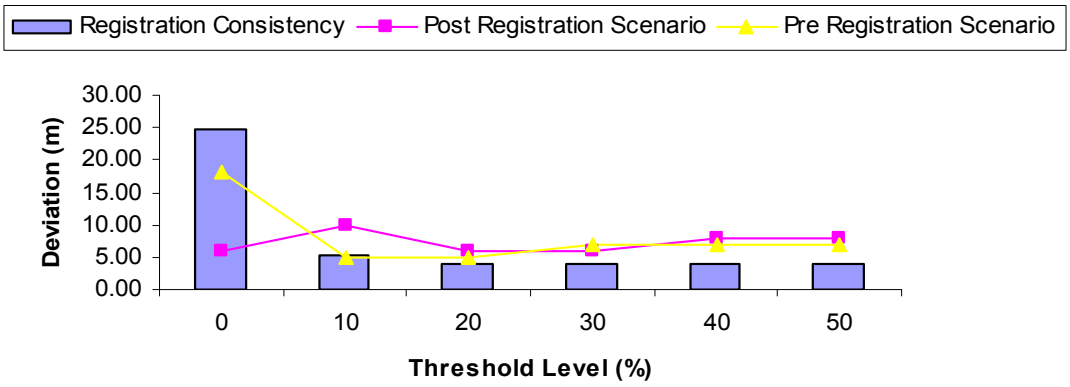


Figure 5.10: MI search spaces generated for dataset-1b. Figures on the left represent the pre registration scenario and on the right represent the post registration scenario. The peaks obtained by mutual information are provided on top right.

(a) 4 m pixel spacing



(b) 2 m pixel spacing



(c) 1 m pixel spacing

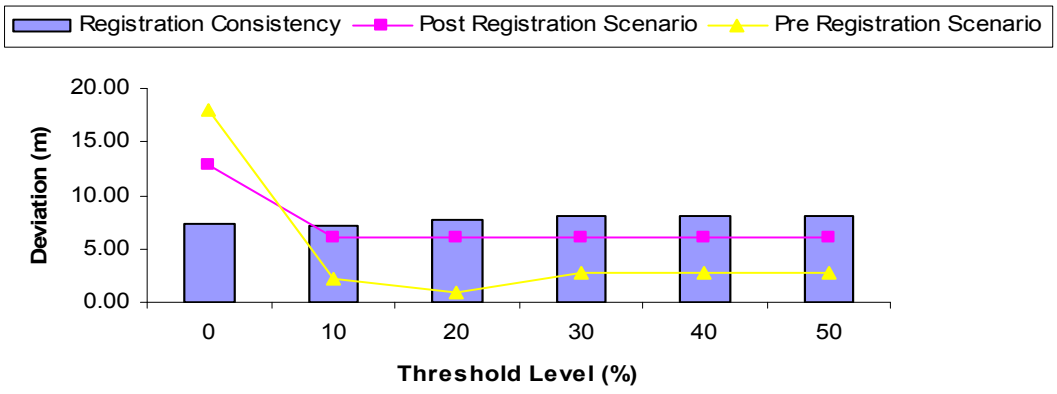


Figure 5.11: Mutual information performance using the thresholding scheme at different pixel spacing for dataset-1b (Figure 5.3). Registration consistency is computed using Equation 5.2 and individual deviation for post and pre registration scenarios are computed using Equation 5.1.

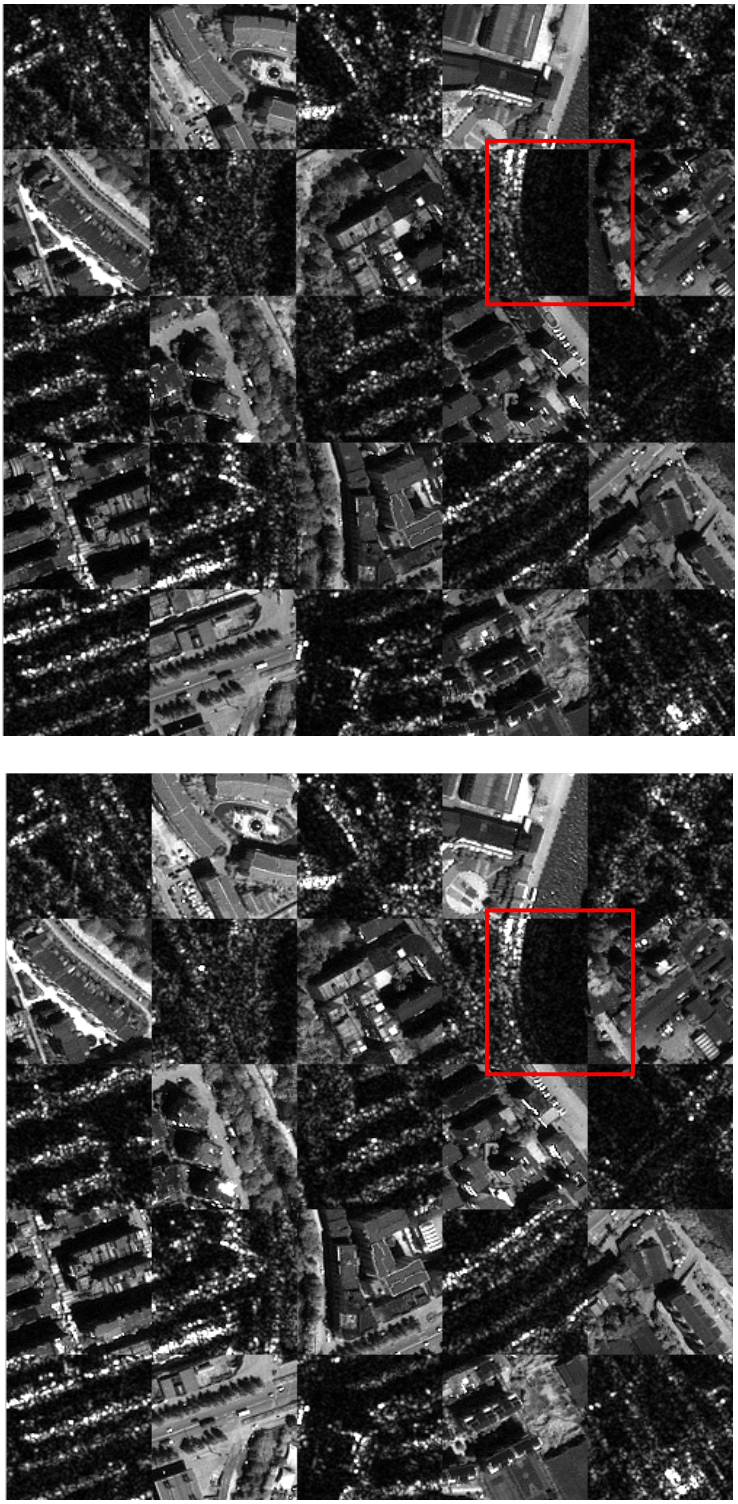


Figure 5.12: TerraSAR-X image (Figure 5.3) check-squared with registered pre disaster IKONOS image. We have utilized (top) before thresholding (-110, -17) and (bottom) after thresholding (-123, -29) parameters. The difference in the river alignment is highlighted in the two images.

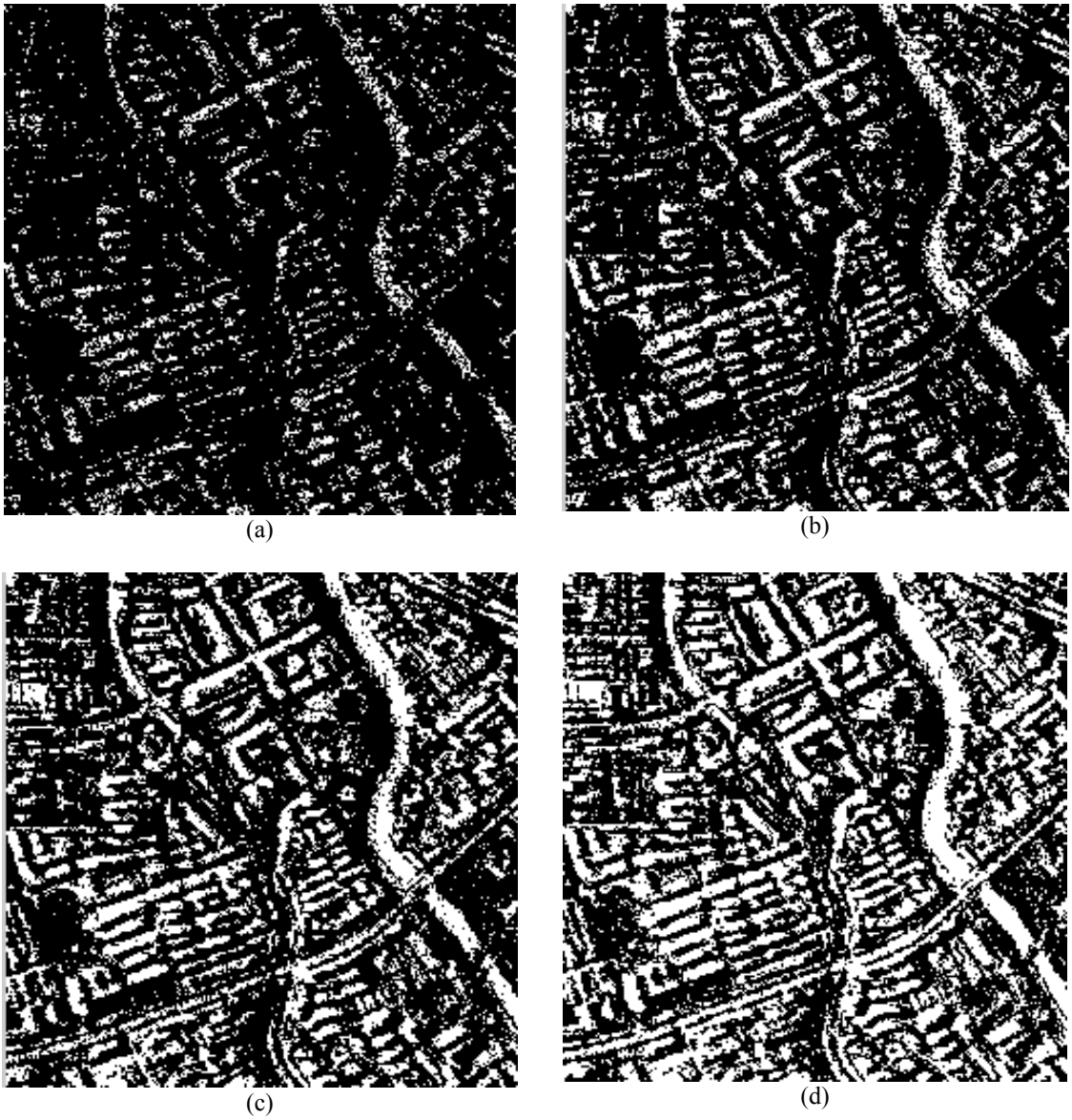


Figure 5.13: Influence of the low thresholds in the range of 10-40% (a-d) introduced on the TerraSAR-X image of dataset-1b (Figure 5.3). It is observed that thresholding the histogram from the lower end of the SAR image takes a lot of pixels away from the "on ground" features like the river and the roads.

It can be anticipated that there is some discrepancy in the introduced thresholding scheme. High end SAR image pixels can definitely remove the bright double bounce pixels but still we have very dark SAR image pixels generated explicitly by Radar shadow and with the current implementation they are participating in the MI statistics. It is comprehended that the removal of dark pixels is not very straight forward as these have a significant representation in dominant features like roads and river in the selected dataset. A sample influence of introducing thresholds

from 10-40% on the TerraSAR-X image of dataset-1b is provided in Figure 5.13. It is observed that along with the shadow pixels most of the on ground feature pixels would be left out from MI statistic computation from such a thresholding scheme. Tests performed on the investigated datasets confirmed this hypothesis as in most of the case the registration peaks do not change and sometimes become more erroneous and/or yield rough search spaces.

In this section, we analysed the MI performance by introducing high end thresholds in the reference SAR image. Introducing high end thresholds in the SAR image produce significant shifts in the MI peaks as most of the rejected pixels in this case are from the SAR double bounce effect. Similar results are not obtained on removing the lower end pixels as these have a mix of on ground feature pixels and Radar shadow. Considering the results from both the datasets with zero overlap, we obtain (-96, -47) and (-94, -50) pixel translation in x and y direction for the post disaster registration and (-122, -30) and (-123, -29) pixel translation in x and y direction for pre disaster scenario. These parameters definitely are erroneous to a certain extent but also depict a consistent mutual information performance after the investigated thresholding scheme. Without an ancillary source to estimating the registration shifts manually between the high resolution IKONOS and TerraSAR-X imagery can be very subjective. It can still be considered that the discussed thresholding scheme may not be adequate to get rid of all the pixels specifically generated by the SAR image geometry for achieving correct registration parameters using MI. Henceforth, in the following section we utilize a region growing technique to again evaluate the MI performance for the same datasets tested here.

5.3.2 Region Based Registration

In this section we present a novel approach for high resolution image registration. This method presents a hybrid methodology combining feature and intensity based approaches for achieving image registration. The chief idea here is to extract high resolution features which are expected to be on ground and following use of their intensity values to estimate registration parameters using mutual information. With this methodology we attempt to highlight the following (Suri et al., 2009b):

i. The usefulness of region growing technique for high resolution feature extraction.
ii. Capability of MI to match images with selected information in only one of the images participating in the registration process.

Considering the very incompatible information generated by the two sensors in urban areas (especially for 3D objects) the idea here is to select only on ground 2D features (image appearance is not greatly influenced by sensor geometry) for registration purposes. In general, for high resolution satellite images acquired over urban areas, common city features like wide roads, rivers, big stadiums, play grounds, parks can be expected to appear in considerable sizes and be represented by relatively homogeneous intensity values (if imaged without occlusions, especially in the case of SAR sensor). This hypothesis gets confirmed as we analyse the appearance of prominent features like river and roads in the images depicted in Figure 5.2 and 5.3. Considering this regular, homogeneous appearance of "on ground" urban features, we intend here to utilize region growing techniques to obtain homogenous features within the images.

The intended registration chain is depicted through a flow chart in Figure 5.14. To initiate the registration process, smoothing operation is mandatory before targeted flat homogeneous regions can be extracted for registration purposes. Considering the image statistics, we employ an enhanced Frost filter (Lopes et al., 1990) for SAR image. The basic Frost filter (Frost et al., 1982) comes into the class of adaptive filters and has been found useful for SAR image speckle reduction. The idea behind smoothing is to increase the homogeneity of the regions to supplement the following important region growing step. Following the region extraction using the standard region growing technique (Castleman, 1996), we attempt to filter the obtained homogeneous regions on basis of their shape properties like area, length and geometrical moments. Followed by the region extraction and filtering based upon geometrical moments we use image morphological operation to obtain the region skeleton. Finally, the extracted skeleton is used for taking original image information (intensity values) from SAR image and is matched with the original optical image using mutual information. We further present and discuss the results obtained from the processing chain illustrated in Figure 5.14 for both the selected datasets.

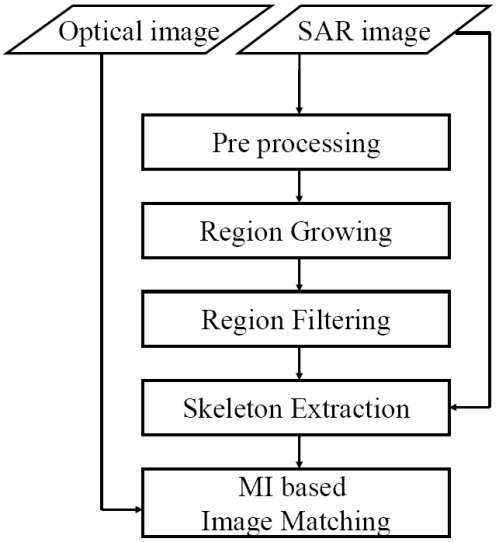


Figure 5.14: Proposed chain for registering high resolution SAR and optical satellite imagery acquired over urban areas.

Results from dataset-1a

The result of the region growing and the filtering step on the reference SAR image for dataset-1a are provided in Figure 5.15. It is visualized that a lot of "on ground" features like the river and the roads are extracted automatically. Following the middle axis extraction using binary morphological operations the skeleton broadened with SAR image information using different window sizes. In this methodology, window size becomes a parameter of evaluation as it is assumed here that the middle axis extracted represents the true middle axis of the desired feature and by selecting various window sizes information surrounding the axis pixels can be accommodated in the MI statistics. Depending upon the width of the features present in the dataset we analyse MI response with respect to a window size range of 5 to 19 m in both the datasets. Keeping the broadened skeleton images (Figure 5.16) as the reference, these have been matched with the input pre and post disaster IKONOS imagery using MI.

As all the extracted features would generally not have the same width and therefore the urban establishments surrounding them (especially for the narrow roads) might start creeping in the MI statistics for certain window sizes. Therefore, we also recommend performing a 10% threshold from the higher end of the SAR skeleton image histogram. This assumption is also in consistence with the observation that most of the ground pixels (especially for SAR images) are situated in

the lower end of the skeleton image histogram. This step in general is found to facilitate much more consistent MI performance in the presented scenario. The results of the matching operation with various window sizes for dataset-1a are tabulated in Table 5.4, sample MI search spaces are provided in Figure 5.17. The search spaces for this region based strategy have been generated using Sextic B-spline kernel in GPVE (Suri and Reinartz, 2008) otherwise the search spaces have been found to be rough due to very less number of pixels participating in the registration process.

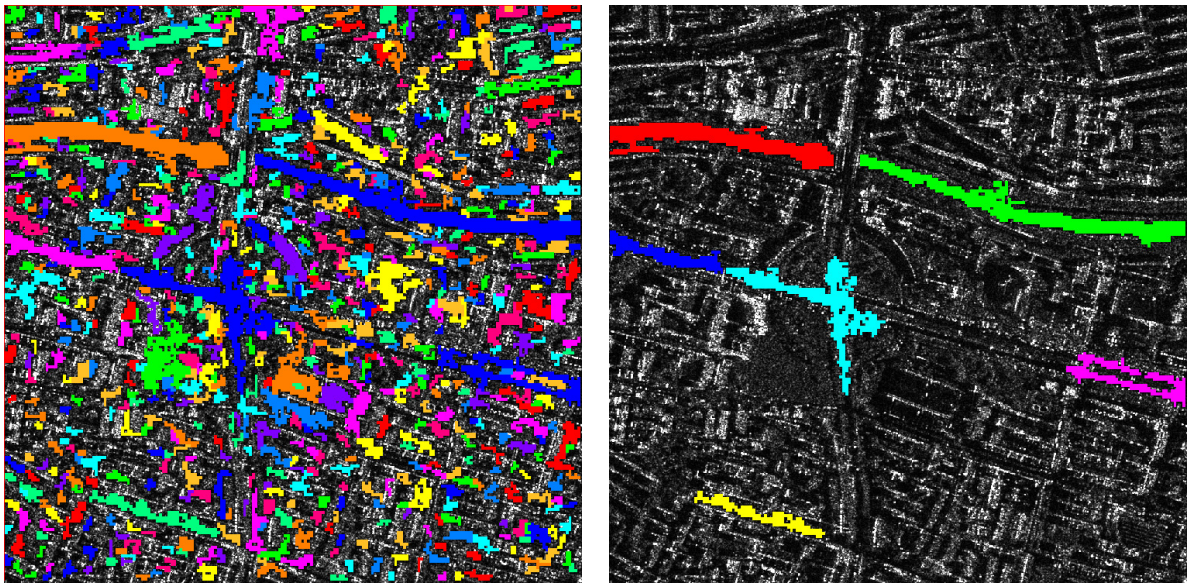
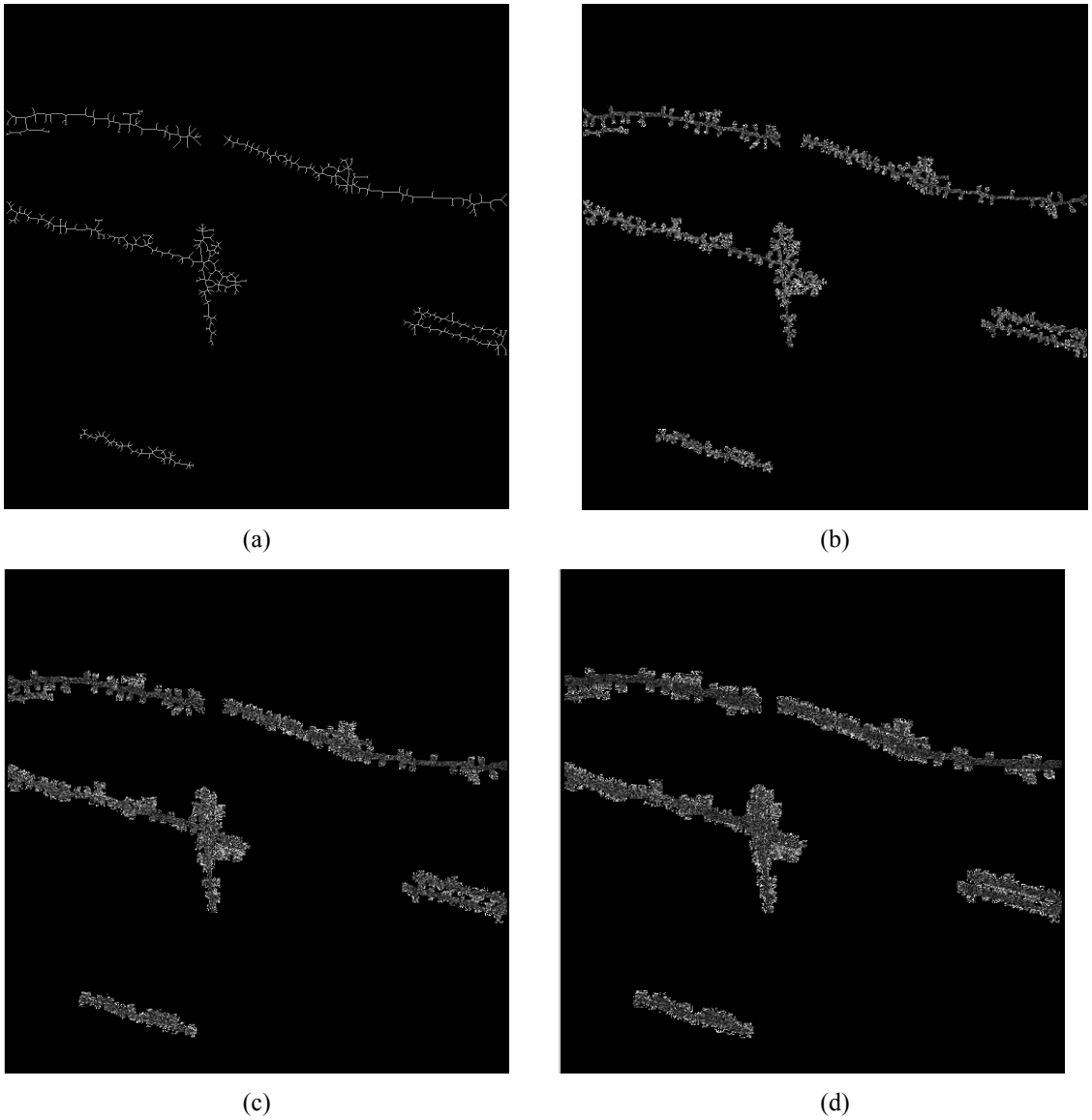


Figure 5.15: Automatically extracted and filtered regions from the TerraSAR-X image in dataset-1a.

For this dataset taking pixels mainly from the "on ground" features has led to a better mutual information performance both in terms of the overall registration accuracy and the individual accuracies of the pre and post registration scenarios (Figure 5.18). A comparison of utilizing MI with and without thresholding and by utilizing the region based scheme with different window sizes for 1 m images can be visualized in Figure 5.18. It is worthwhile to mention that by just selecting desired features in the SAR image has led to a deviation (Equation 5.1) around 3 m in the pre registration scenario and around 6 m for the post registration scenario. The same dataset with the global thresholding scheme produced respective deviations of 5 and 10 m respectively. We further examine this registration scheme for dataset-1b to complete the evaluation.

Table 5.4: Registration results for dataset-1a (Figure 5.2) using region based scheme at 1 m pixel spacing.

Window Size	Post IKONOS to TerraSAR-X	Pre IKONOS to TerraSAR-X	Registration Consistency
5 m	-85, -44	-120, -24	2.00
7 m	-87, -44	-119, -28	6.40
9 m	-86, -44	-120, -29	5.83
11 m	-87, -44	-124, -30	6.00
13 m	-88, -44	-124, -30	6.08
15 m	-87, -45	-123, -30	5.10
17 m	-94, -46	-130, -32	6.10
19 m	-91, -47	-123, -30	5.83



(a) (b)

(c) (d)

Figure 5.16: (a) Extracted skeleton from the regions obtained in Figure 5.15. (b-d) skeleton broadened with original SAR image intensity values (Figure 5.2a) using a window size of 7, 11 and 15 meter respectively.

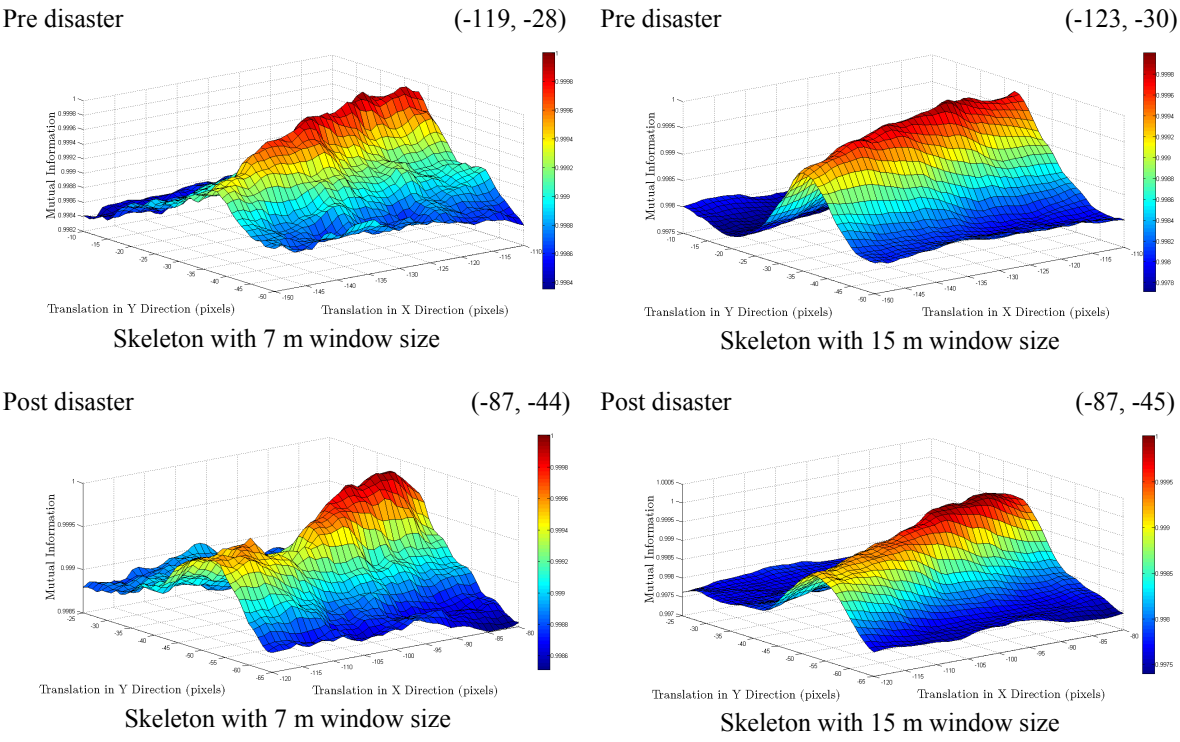


Figure 5.17: MI search spaces generated for pre and post disaster registration scenarios using skeleton generated by window size of 7 and 15 meter.

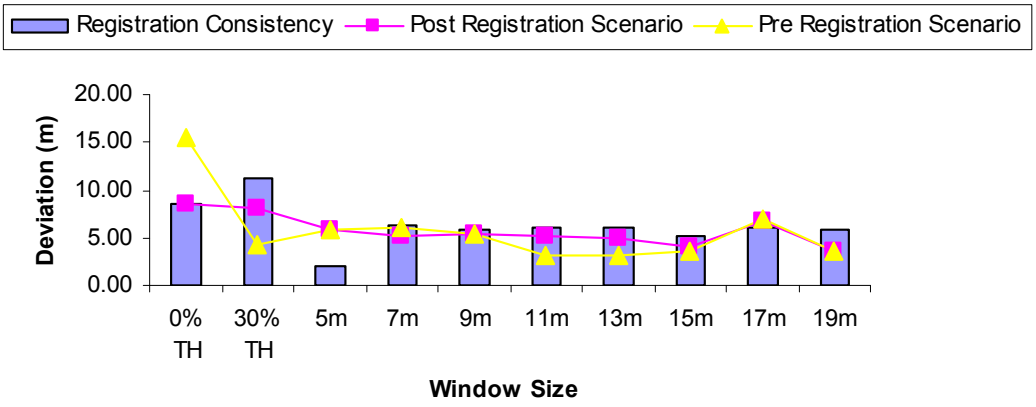
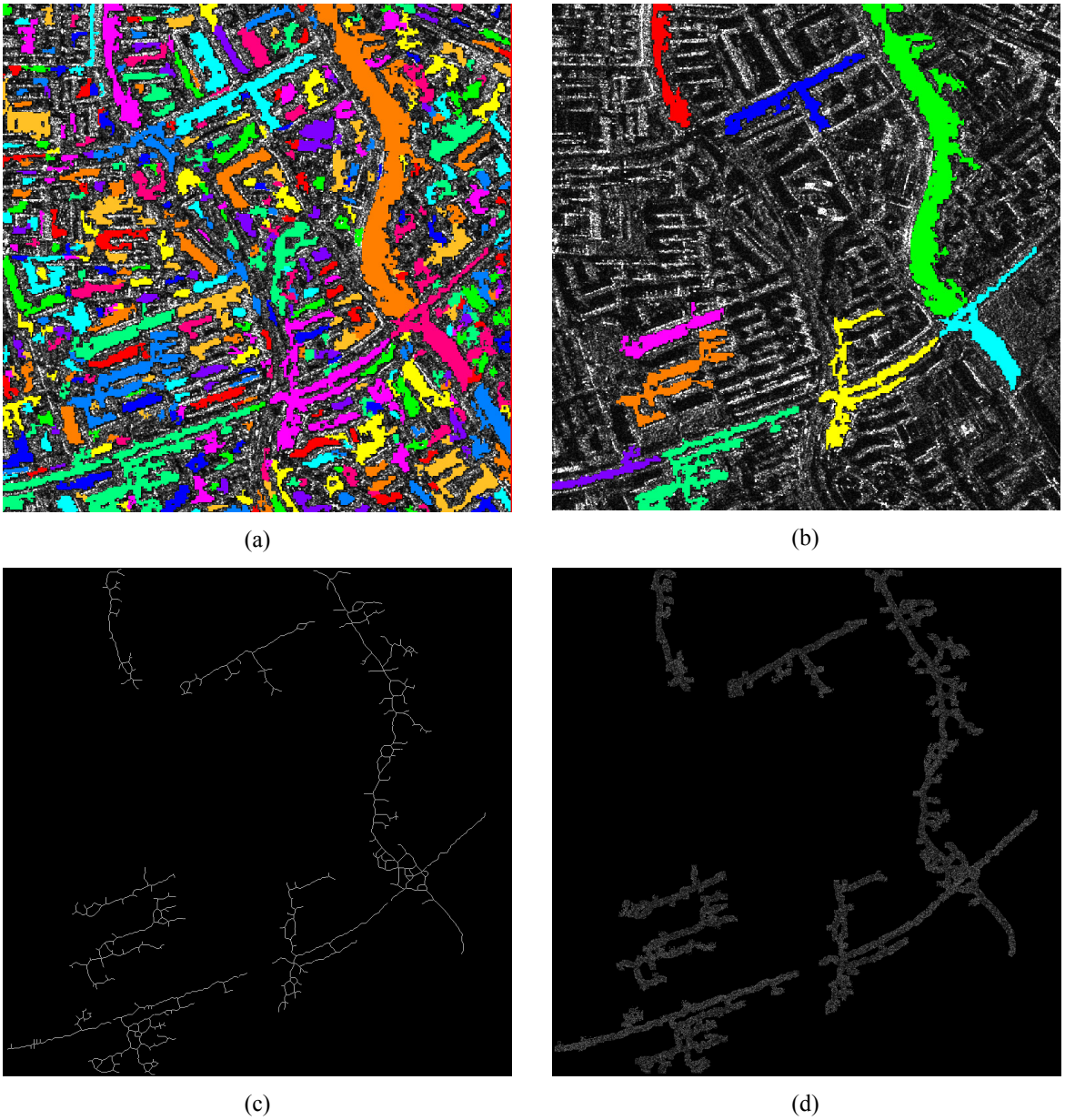


Figure 5.18: Mutual information performance using the region based registration scheme for dataset-1a. Here we compare the application of MI with and without thresholding (0 and 30%) and after utilizing extracted "on ground" regions with different window sizes with 10% thresholding. Registration consistency is computed using Equation 5.2 and individual deviation for post and pre registration scenarios are computed using Equation 5.1.



(a) (b)

(c) (d)

Figure 5.19: (a) Extracted regions from the TerraSAR-X image of dataset-1b (Suri et al., 2009b) (b) Filtered regions. (c) Obtained skeleton using image morphological operations. (d) Skeleton broadened with original SAR image intensity values using a window size of 11 meter.

Results from dataset-1b

The results of the region growing and finally the skeleton extraction step on the reference SAR image for dataset-1b are provided in Figure 5.19. The results of the matching operation are tabulated in Table 5.5, MI search spaces provided in Figure 5.20. For dataset-1b, the region based scheme produces an average deviation of less than 5 m for both the registration scenarios

where as the global thresholding scheme produced a very similar performance yielding 6 m for the post disaster and approximately 3 m deviation for the pre disaster registration scenario (Figure 5.21).

Table 5.5: Registration results for dataset-1b (Figure 5.3) using region based scheme at 1 m pixel spacing.

Window Size	Post IKONOS to TerraSAR-X	Pre IKONOS to TerraSAR-X	Registration Consistency
5 m	-89, -50	-126, -25	5.00
7 m	-89, -50	-125, -29	1.41
9 m	-86, -44	125, -28	3.16
11 m	-91, -50	-126, -26	6.32
13 m	-90, -52	-127, -27	5.00
15 m	-91, -52	-127, -27	5.10
17 m	-91, -52	-127, -29	3.16
19 m	-91, -52	-127, -29	3.16

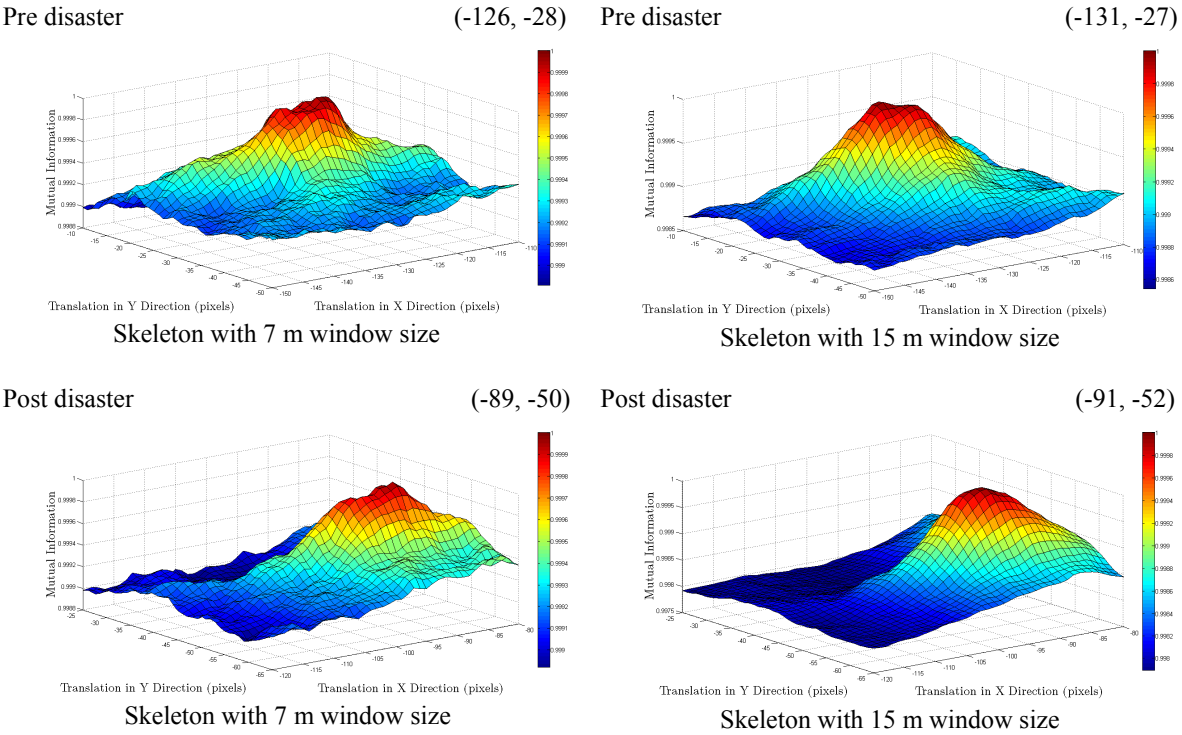


Figure 5.20: MI search spaces generated for pre and post disaster registration scenarios using skeleton generated by window size of 7 and 15 meter.

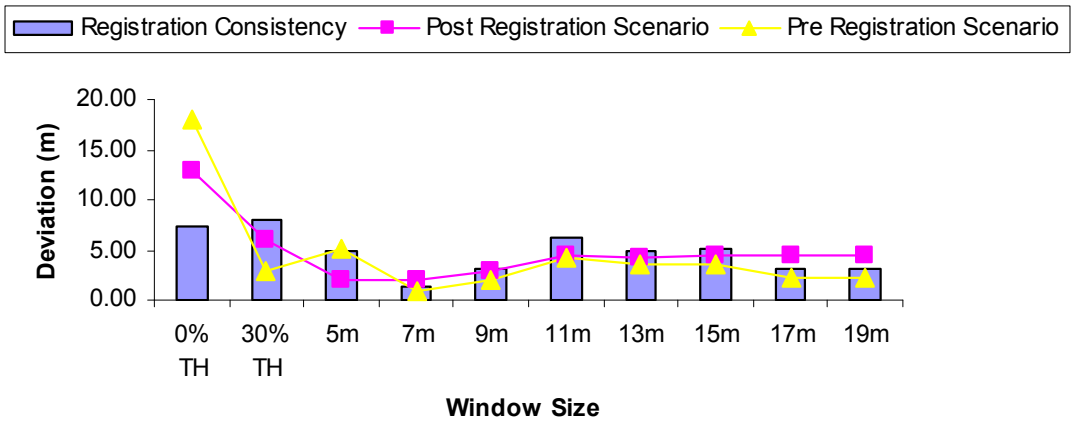


Figure 5.21: Mutual information performance using the region based registration scheme for dataset-1b. Here we compare the application of MI with and without thresholding (0 and 30%) and after utilizing extracted "on ground" regions with different window sizes with 10% thresholding. Registration consistency is computed using Equation 5.2 and individual deviation for post and pre registration scenarios are computed using Equation 5.1.

5.3.3 Outlook Based on Obtained Results

On the basis of underlying similarity metric we have presented two distinct methods for achieving automatic registration between high resolution SAR and optical images. To summarize, the first method is based upon histogram thresholds introduced in the reference SAR image and the second method is based upon certain features extracted again in the reference SAR image which are expected to be imaged independent of any sensor geometric configuration. In both the methods, the optical image has not been processed in any ways and good registration results have been achieved. We can compare the above methods on the basis of the following criteria:

Consistency: It can be realistically argued that the region based scheme has produced better registration results than the global thresholding scheme. This can mainly be attributed to the fact that maximum of the pixels contributing to the joint histogram in this case were actually "on ground" pixels thus facilitating an improved MI performance. On the other hand, the global thresholding scheme seems to be slightly inaccurate due to the discrepancy introduced by the thresholds. Pre-dominantly double bounce pixels are removed from the joint histogram but still lot of the dark pixels generated specifically by geometrical characteristic of the SAR sensor like

Radar shadow participate in the registration process which can make the entire process slightly inaccurate.

Applicability: Considering the methods discussed, the region based registration method looks less likely to have a very general applicability. As extraction of "on ground" features like roads, rivers, play grounds may not always be possible automatically and this can jeopardise the entire process. Proposed region based scheme has also been found to be effective for relatively large scenes with dominant features which might not always be present. On the other hand the histogram based thresholding method can be utilized for more general registration purposes as it involves a subtle feature extraction step which can be easily realized using the SAR image histogram. Practically, the region based registration scheme can also be employed effectively by using a classification scheme to extract "on ground" feature pixels or by including ancillary information for feature extraction (like road databases to extract road skeletons) in the registration process.

To further evaluate and validate the thresholding based image matching method in the following section we present a novel application of SAR optical image matching to improve the sensor orientation and orthorectification of high resolution optical data.

5.4 IMPROVEMENT OF SENSOR ORIENTATION AND ORTHORECTIFICATION OF HIGH RESOLUTION OPTICAL DATA

Until now, we have shown the capability of MI to match multimodal SAR and optical images. Normally, matching has been implemented as a problem of determining shifts in x and y direction between two images of interest. Simply determined x and y directional shifts might well serve the purpose for small image chips but are not sufficient to register large remote sensing images. This conventional application of MI was shown to suffer (see Section 3.4.4) where we analysed MI performance for Landsat and Radarsat (approx 3000x3000 pixels in size). It was highlighted that simply determining the shifts yielded huge local distortions within manually marked registration tie points. There we proposed to use MI locally within the images to approximate an affine transformation between already orthorectified images to reduce the

local distortions. This technique might suffice for sub scenes but would definitely not meet the desired accuracy requirements for complete coverages produced by modern high resolution satellites.

The idea we investigate in this section is to utilize MI locally within the images and then utilize the obtained matching results to improve the exterior orientation (Müller et al., 2007) and thereby improving the geometric accuracy of optical satellite data (Reinartz et al., 2009). The main motivation of this exercise can be summarized as follows:

i. The geometrical accuracy of the new German satellite TerraSAR-X has been found to be within few meters for the Enhanced Ellipsoid Corrected (EEC) images (Nonanka et al., 2008). The Single Look Slant Range Complex (SSC) mode has a better accuracy than 1 m and this accuracy levels are maintained if the data are orthorectified incorporating a high precise DEM.

ii. The accuracy of orthorectified high resolution optical satellite data suffers due to insufficient knowledge of satellite attitude and thermal influenced mounting angles. Errors in this scenario have been found in the range five to several hundred meters (depending upon satellite configuration).

iii. The accuracy level of optical orthorectified imagery can be improved by precise ground control. Source of this ground control information can be GPS measurements, topographic maps or very accurate reference images (Jacobson, 2004). In case we have access to accurate optical reference data we can utilize completely automated and robust optical matching techniques (e.g. least squares matching) to achieve precise ground control. This technique of using accurate optical reference images has been shown to be very precise with an overall accuracy of around half pixel size (Müller et al., 2007). However, the main limitation in the above process is the lack of precise enough reference images especially for high resolution data.

iv. Considering encouraging results discussed from the application of MI to high resolution imagery in previous section and the availability of accurate images from TerraSAR-X, it looks feasible to extract accurate ground control (through local image matching) by using TerraSAR-X images as the reference source.

By doing this exercise we can also quantify the absolute registration accuracy (available ground truth) achieved by MI for metric resolution images acquired by multimodal sensors using different geometries, which has not been found very abundantly in the reviewed literature. Specifically, we concentrate on TerraSAR-X and IKONOS imagery, which have been the topic of discussion so far in this chapter.

5.4.1 Processing Chain

Considering the application demands, we test the histogram based registration technique presented in Section 5.3.1. The dataset under consideration comprises of images acquired over the city of Munich, Germany. The details of the dataset are tabulated in Table 5.6. The images have an acquisition difference of three years and mostly have an urban land cover.

Table 5.6: Details of the imagery utilized for improving optical sensor orientation and orthorectification (images acquired over the city of Munich).

	TerraSAR-X	**IKONOS-2**
Mode	High resolution spot light (HS) Descending orbit	Forward scanning
Spectral Resolution	9.65 GHz	450 - 900 nm
Spatial Resolution	Ground range: 0.96m Azimuth: 1.12 m	Cross Scan: 0.83 m Along Scan: 0.84 m
Geometry	Incidence Angle: 39.3º	Sun azimuth: 153.9 º Sun elevation: 61.5º Collection azimuth: 354.9º Collection elevation: 80.8º
Bits per pixel	16	11
Date of Acquisition	30/09/08	15/07/05
Product	EEC product	IKONOS GEO

The processing chain utilized in this section is depicted in Figure 5.22. We initiate by forming a grid of conjugate points in both the images which are to be fine matched using MI

(Figure 5.23). As the images are already geocoded we can place this regular grid using the initial geocoding information of both the TerraSAR-X and the IKONOS image. Grids have to be placed so that every point gets a decent window size to be utilized for computing the mutual information statistics from both the images. Further, we have used the following parameter settings for all the chips to be matched:

i. The local MI based matching is carried out using the histogram based registration scheme presented in Section 5.3.1. We threshold each of the reference TerraSAR-X chips with 20 percent high threshold and also 10 percent low threshold locally and match them with the input IKONOS chips.
ii. As for small image chip sizes, MI surfaces are generally found to be rough and suffer from interpolation artefacts, we have selected GPVE technique using Sextic B-spline kernel for estimating the joint histograms of bin size 100.
iii. Multilayer image chip matching forming two layers in each of the chips is utilized.
iv. For optimization, we utilize the SPSA optimization scheme as described in Section 4.4.

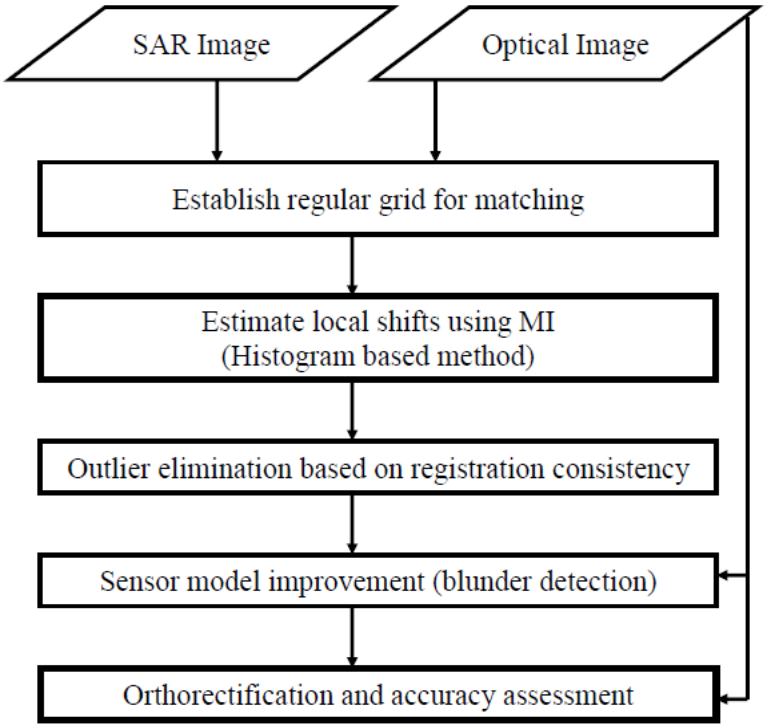


Figure 5.22: Processing chain for sensor orientation and orthorectification of IKONOS data.

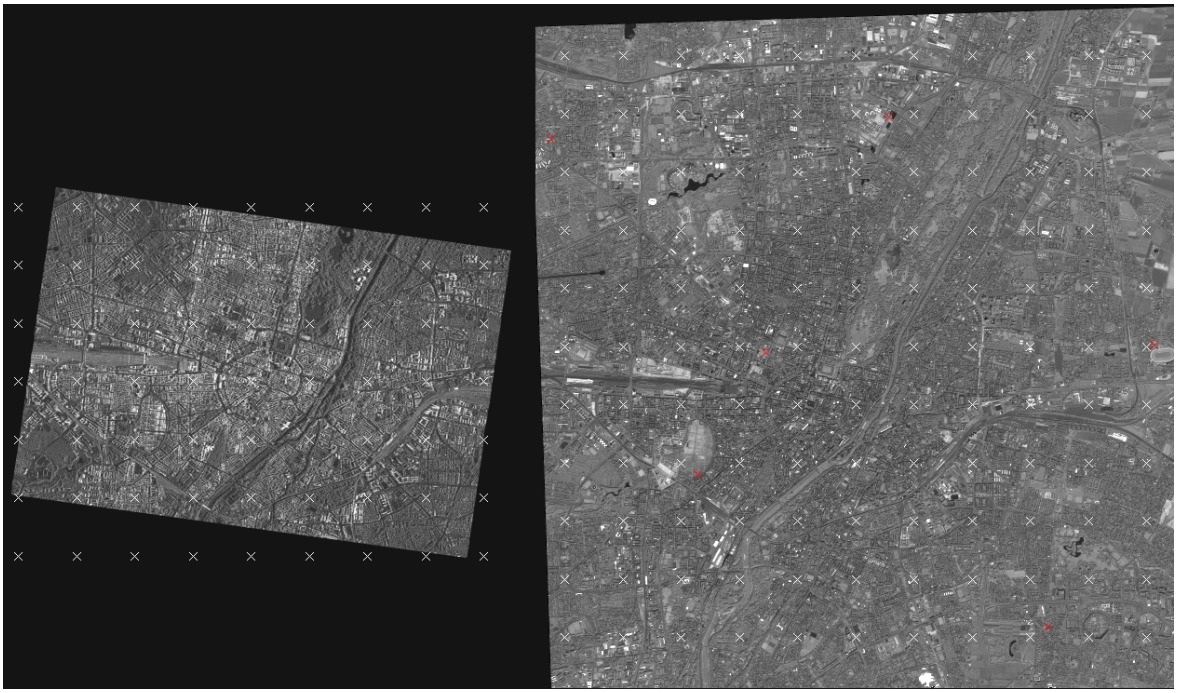


Figure 5.23: Established regular grids for local matching between TerraSAR-X (left) and IKONOS (right) images. The White crosses represent the points used for local matching using MI and the Red crosses represent the check points used later to judge the quality of orthorectification.

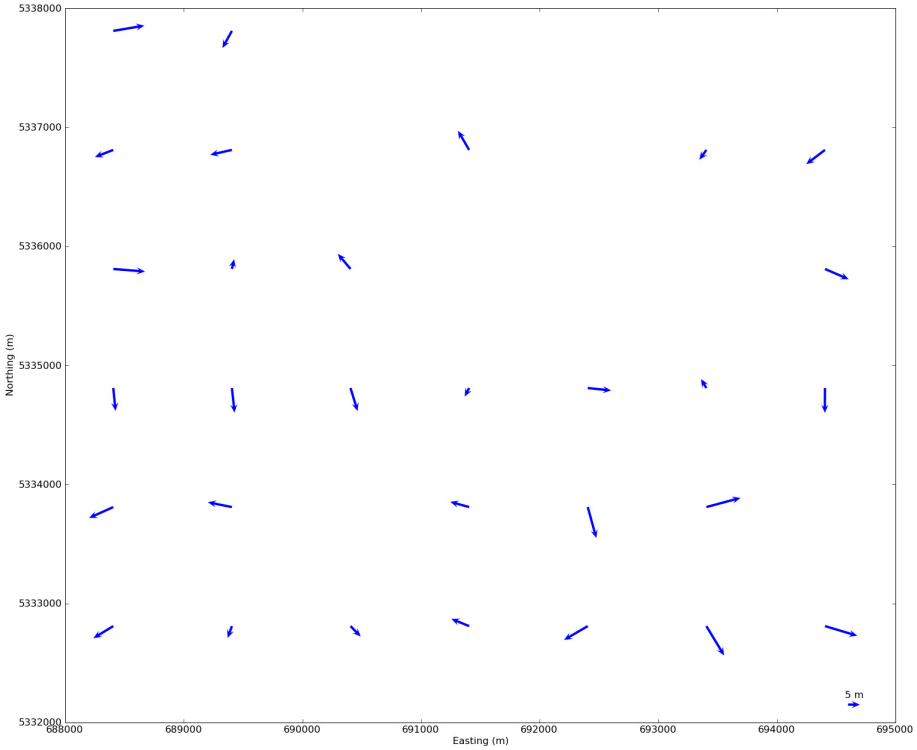


Figure 5.24: Local shifts obtained for IKONOS grid points as estimated by MI using TerraSAR-X image as the reference (no segmentation).

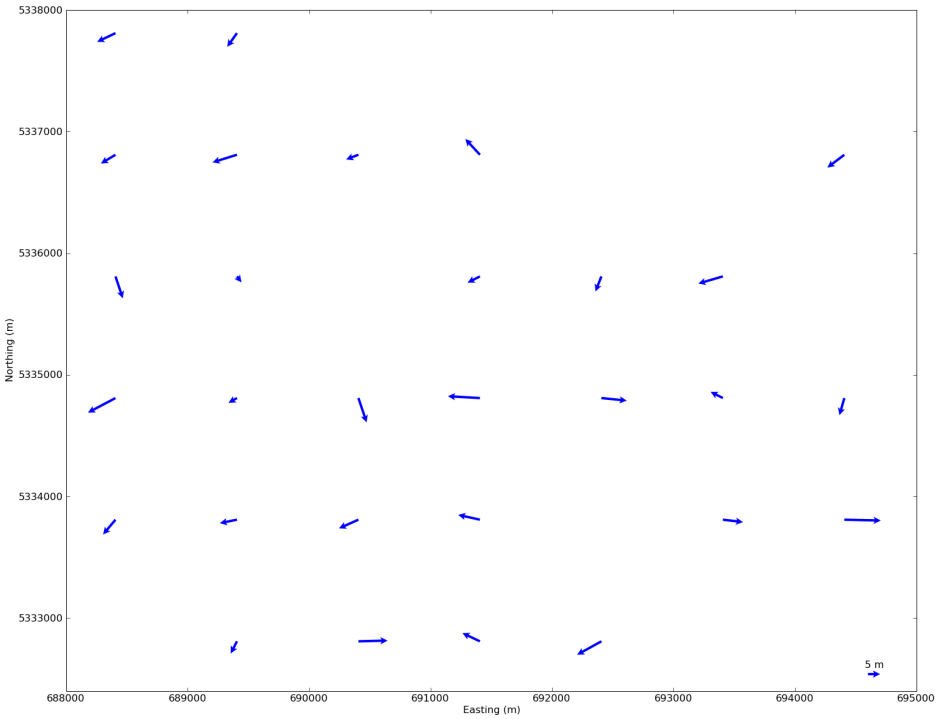


Figure 5.25: Local shifts obtained for IKONOS grid points as estimated by MI using segmented TerraSAR-X image as the reference.

We perform a two way matching, first keeping TerraSAR-X as reference and IKONOS-2 as input chips and then we revert the settings to obtain matching consistency (see Section 2.4.3 and 3.4.4) which as earlier helps us to remove some of the bad matches (according to a set threshold). As the regular grid is just placed without any marking of favourable high entropy/contrast areas like crossings, intersections, runways there would be points in very homogeneous regions where MI matches generally tend to fail. Therefore, performing a two way optimization and checking for match consistency helps in removing some of the wrong matches.

Complementing the tie points found from the TerraSAR-X data with interpolated height values from the DEM leads to 3D GCPs (in case DEM is not available we use a constant height) in the image coordinates of the IKONOS data, which are utilized to improve the sensor model of the optical system (here RPC based). A method for blunder detection is integrated in the least squares adjustment, which eliminates iteratively GCPs with a residual greater than a threshold starting with the most deviating GCP. In this context a residual is defined as deviation of GCP coordinates from the re-calculated object point coordinates using the refined sensor model,

which is derived from the adjustment using all currently valid GCPs. In a first step the whole GCP set serves as input for the iterative parameter estimation. Successively the GCP with the worst deviation and residuals greater than a threshold is removed from the GCP set. This iterative procedure is repeated until a GCP set remains, which is consistent to the sensor model. The procedure has to be iterative, because each GCP influences the result of the parameter estimation. As threshold for the GCP blunder detection two pixel sizes has been found to be sufficient for ALOS / PRISM and IKONOS images (other values like 1 pixel size threshold show no significant improvement of blunder detection). Although this process is deterministic, the drawback of this method is that systematic errors for the majority of the GCPs can lead to erroneous or shifted values of the estimated parameters. Another possibility is the fact that we might obtain a poor distribution of the GCPs concentrated in only one part of the scene and a model estimated from such a distribution might not be ideal and consistent for the entire coverage.After obtaining the refined sensor model using the locally matched grid points, we orthorectify the IKONOS image and check the results both before and after the improved sensor orientation with well spread out reference coordinates extracted from a much accurate reference data[12] (expected accuracy of about 2 m).

5.4.2 Results of Orthorectification

The process starts with local image matching between the reference TerraSAR-X and IKONOS image using MI. For a comparison of the results, we perform matching both with and without the proposed histogram based method. The initial shift obtained within the established 121 grid points with and without introducing SAR image thresholds can be visualized in Figure 5.24 and 5.25 respectively. Due to various thresholds like registration consistency, image background and image chip entropy introduced we obtain MI estimated shifts for only 29 points (with segmented TerraSAR-X image in Figure 5.25), rest of the 92 point pairs get rejected due to above mentioned reasons.

[12] www.geodatenzentrum.de

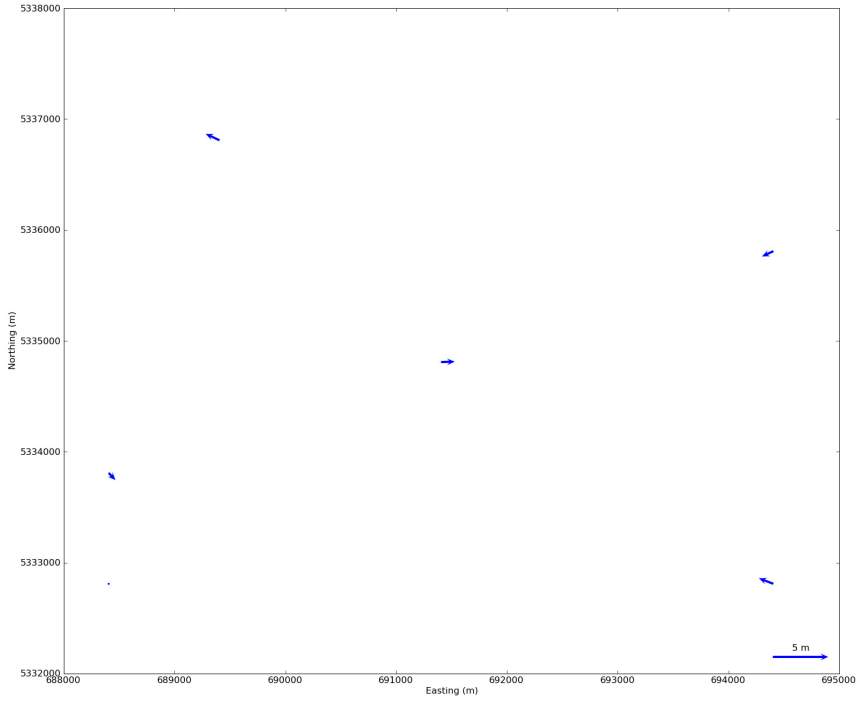


Figure 5.26: Residual vectors obtained after using TerraSAR-X image as the reference (6 points after outlier elimination contributed towards sensor model correction).

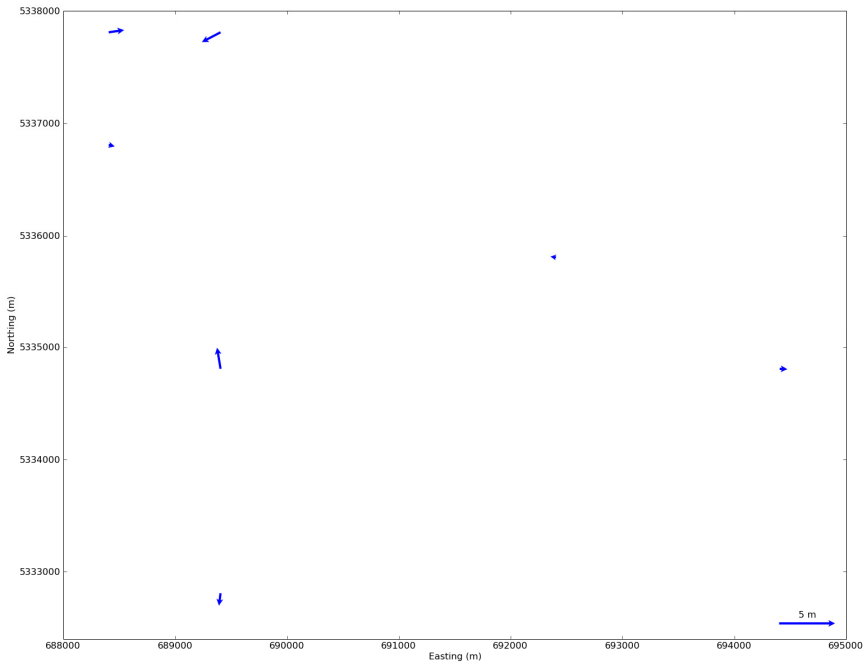


Figure 5.27: Residual vectors obtained after using segmented TerraSAR-X image as the reference (7 points after outlier elimination contributed towards sensor model correction). This set of seven points does not reflect the ideal desirable distribution of control points for accurate orthorectification.

For the case, where we don't introduce any segmentation we obtain shifts within 31 grid points (Figure 5.24). In both the cases, the shifts are obtained using the parameter settings mentioned in the previous section.

Once these shifts are obtained we use the TerraSAR-X geocoordinates and IKONOS image coordinates with a constant elevation height to correct the RPC sensor model of IKONOS-2. The scheme further performs blunder detection and then selects a subset of matched points (acting as GCPs) to do the desired task. The residuals obtained from the estimated shift files for the two MI matching schemes can be visualized in Figure 5.26 without segmentation and Figure 5.27 with segmentation. For RPC correction, we finally obtain seven (with segmentation) and six (without segmentation) points for sensor model correction from both the matching schemes. Generally, it has been found that we obtain a higher percentage of points (especially for scenes having mainly rural or sub urban settlements) for sensor model improvement but that has not been the case for this dataset. This can mainly be attributed to entire urban land cover present and utilized for image matching. The residuals estimated by the built sensor model for the two cases are provided in Figure 5.26 and Figure 5.27.

Table 5.7: Shift in meters obtained at six control points after different orthorectification schemes.

	TerraSAR-X not utilized to establish ground control		Ground control extracted from TerraSAR-X using simple MI based matching		Ground control extracted from TerraSAR-X using segmentation prior to MI based matching	
Point Number	X	Y	X	Y	X	Y
1	4	-8	-4	2	-4	-5
2	4	-7	17	7	0	3
3	5	-8	2	-5	1	-3
4	4	-9	24	-4	5	0
5	5	-9	-5	-11	2	-5
6	7	-12	8	-19	7	-5
Average (m)	4.83	8.83	10.00	8.00	3.17	3.15
Std Dev (m)	1.72	1.72	11.66	9.23	3.87	3.33

After obtaining the correct sensor model we orthorectify the IKONOS-2 images using the two models and compare the results with 6 ground control points (GCP), with their location on the IKONOS image shown in Figure 5.20. The residuals obtained at the 6 GCPs for the initial orthorectification (before sensor model improvement), orthorectification through ground control extracted from MI based local chip matching both with and without segmentation of the TerraSAR-X image chips are tabulated in Table 5.7.

It is interesting to note that prior sensor model improvement the initial orthorectification had almost constant shifts (not using TerraSAR-X for extracting ground control) at the six ground control points; this should not be taken as a general trend as we have observed huge local variations for some other scenes tested. Further, it is observed that using TerraSAR-X chips as the reference and not doing any segmentation produces the worst results in the three test cases.

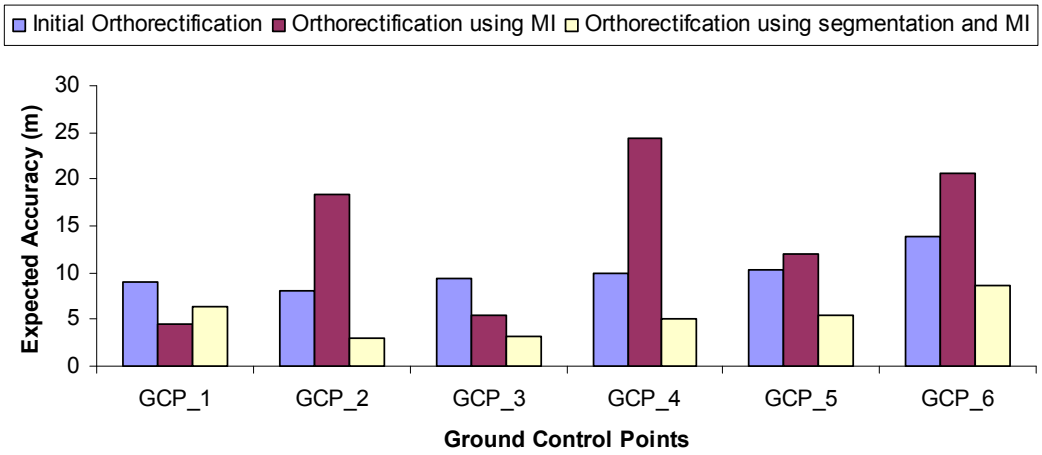


Figure 5.28: Plot of the absolute accuracy achieved by different orthorectification schemes at six reference points.

The best results are obtained in using the proposed histogram based thresholds prior to MI based local image matching. On average we obtain around 6 m absolute accuracy at the six GCP locations, individual absolute errors in the three test cases at the GCP locations can be visualized in Figure 5.28. Further, visualization to compare the three test cases is provided in Figure 5.29, which shows a window of 50m around the six ground control locations and the results of the other three orthorectified IKONOS images.

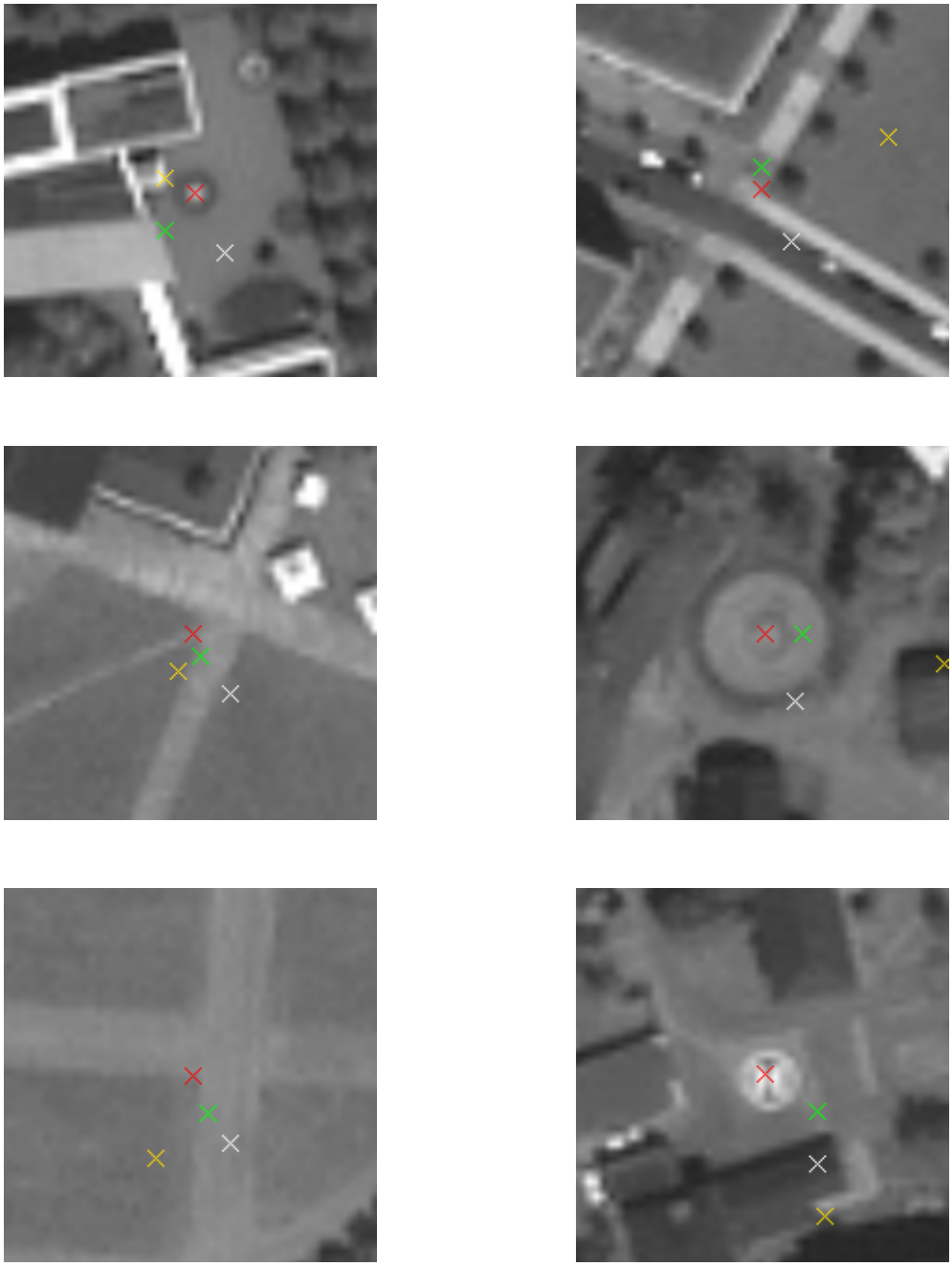


Figure 5.29: Accuracy of different orthorectifation schemes for the IKONOS image at 6 reference points. Red cross indicates the reference ground truth, White cross indicates the initial orthorectification prior to improved sensor orientation, Green and Yellow crosses represent the results of orthorectification after improved sensor orientation using ground control established by mutual information matching with and without segmentation respectively. The image chips represent a 50m window size around the reference points (IKONOS images displayed with 4 times zoom).

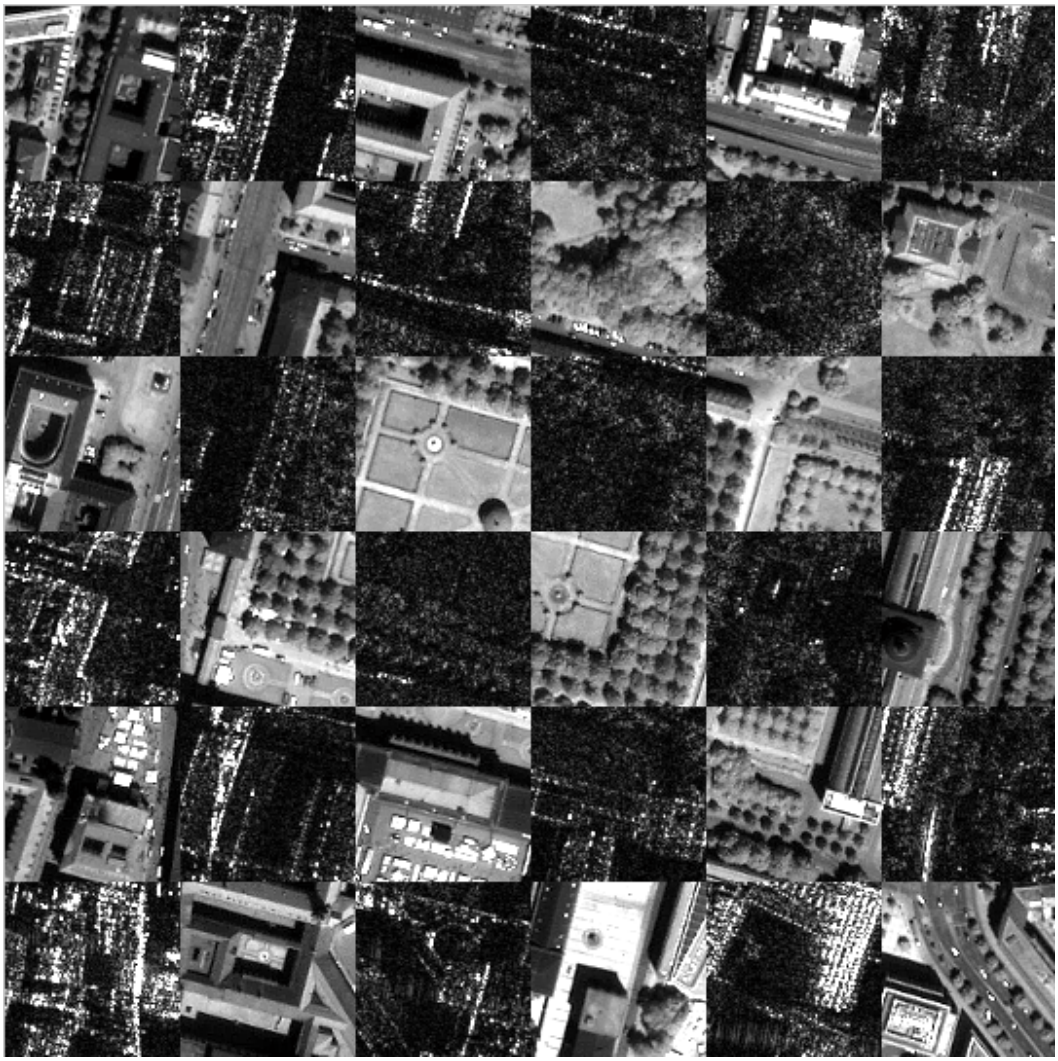


Figure 5.30: TerraSAR-X and orthorectified IKONOS through improved sensor orientation check squared to highlight good registration of "on ground" features like the park at the centre and the adjacent roads. Image area is of size 540×540 pixels and is displayed without any zoom.

It is clear from the discussed process that most of the pixels participating in the registration scheme are intended to be from the on ground features (roads, parks, river etc.) and thus they are expected to show good registration in both the images. For this case of TerraSAR-X and IKONOS image (orthorectified through improved sensor orientation) well aligned "on ground" features can be visualized in Figure 5.30 and Figure 5.31. Especially in Figure 5.31 we also highlight the misregistration much easily observed in case of flat roof buildings as compared to gabled roof tops mainly due to much compatible optical and SAR sensor response produced by the former.

Figure 5.31: Check squared TerraSAR-X and IKONOS image orthorectified through improved sensor orientation. As in Figure 5.30 we observe good registration of "on ground" features but the highlighted flat roof building shows a misalignment. Image area is of size 600x600 pixels and is displayed without any zoom.

5.4.3 Discussion

In this section, we have presented a novel method for sensor orientation and orthorectification of high resolution optical data. The segmentation scheme proposed earlier in this chapter has been found useful to improve MI performance for matching high resolution imagery acquired over dense urban areas. For this dataset acquired over the city of Munich, we obtain around 4.5 m

absolute accuracy (reference data approximately 2 m accurate), in general for several other IKONOS scenes tested with this proposed processing chain has yielded absolute accuracy in the range of 2-7 m. It is commendable that the entire processing chain is completely automatic and its performance has not been found very susceptible to different parameter settings. There are various factors that contribute to the overall accuracy like the actual scene conditions and mutual information matching performance which might not be up to the mark if we have large homogeneous regions in most of the images participating in the matching process.

Further, it is mandatory to have a good distribution of ground control which is utilized for improving the sensor orientation, the distribution obtained for the Munich scene is still not the ideal one. This is an important factor and a failure in achieving this could yield a scenario that we might achieve good registration in parts of the scene and severe misalignment in the other parts. Further, this misalignment might be difficult to observe especially in dense urban areas and in absence of a reference point (for accuracy analysis) in that particular part of the scene. Similarly, for an accurate analysis of the results, a good distribution of check points is very much required as done for this dataset. In general, the MI based ground control establishment looks promising and has given us useful results for orthorectification of IKONOS and ALOS sensor images. (Sample results for ALOS imagery are available in Reinartz et al. (2009)). Generally, the performance has been at its best for rural or sub urban land covers with not too much of 3D objects with very incompatible sensor response. For such scenes, we have almost achieved 1-2 pixel absolute accuracy for IKONOS and ALOS data. Scenes, with very dense urban areas like the one presented in this section give a slightly deteriorated performance with an average accuracy of around 4-8 m for IKONOS data.

5.5 CONCLUSION

This chapter started with a discussion on the challenges encountered in registering high resolution imagery especially acquired over dense urban areas. On the basis of proven capability of mutual information for handling SAR-optical registration scenarios, we have presented two approaches for registration of high resolution imagery in urban areas. One technique is based on the reference image histogram and the other technique requires “on ground” features to be

extracted and then finally matched with the input image. In both the presented methods the pre-processing steps are only required in the reference SAR image. The technique based on reference image histogram looks to be less accurate than the region based technique, but definitely has a much more general applicability than its latter counterpart. On the basis of the histogram based technique, we have demonstrated a completely automated method to obtain ground control from accurate TerraSAR-X images for improving the sensor orientation (RPC's) of IKONOS data. This completely automated technique can be applied for different high resolution optical sensors in the same league of spatial resolution.

CHAPTER 6

SIFT BASED MULTIMODAL SAR IMAGE REGISTRATION

6.1 INTRODUCTION

With this last chapter, we turn our attention to another important objective of this dissertation. Here, we shift our area of interest from SAR-optical registration scenario to multimodal SAR image matching/registration which as pointed out earlier in Chapters 1 and 2, till date has no established supreme methodology. SAR imagery along with its weather independent capability is useful for diverse applications like DEM generation (Dupont et al., 1997), image fusion (Moghaddam et al., 2002), soil moisture estimation (Hegarat-Mascle et al., 2002), traffic related studies (Palubinskas et al., 2005), change detection (Bovolo and Bruzzone, 2005) and many more (Eineder et al., 2005). Considering challenges in feature based multimodal SAR image registration (see Section 2.6) and advancements in the field of computer vision, improvements in point feature matching for applications like radargrammetry (Tupin and Nicolas, 2002) and geometrical SAR image registration look feasible. Henceforth, in this chapter, we evaluate the application of the SIFT (Scale Invariant Feature Transform) operator for remote sensing SAR image matching and registration applications. A detailed conceptual description of the SIFT operator is provided in Appendix of this dissertation, here we discuss SIFT operator and its application in context of SAR images and propose refinements for the same.

6.2 SIFT IN CONTEXT OF SAR IMAGERY

The motivation for considering the SIFT operator for SAR image matching application is the operators scale invariance along with its robust feature identification and matching capabilities. For SAR images, the SIFT operator has the inherent ability to detect and match stable features at lower scales, where the speckle influence is expected to diminish. Its robustness in the face of occlusion, clutter and noise (Lowe, 2004), further facilitates its application for SAR image

matching. Broadly the SIFT processing chain (see Appendix) can be decomposed into three main categories, namely feature detection, identification (rotation and scale invariant descriptor formation) and matching. All the three following components have been developed and optimized for optical images, henceforth we propose and evaluate some changes in the SIFT processing chain with an objective of successfully adapting it for multisensor SAR image matching/registration applications.

6.2.1 Feature Detection

The feature detection is primarily based upon image smoothing by the Gaussian filter followed by the DOG technique to form the scale space, followed by extremum detection (potential features). Further, the sub-pixel feature localization is dependent upon gradients approximated through simple differencing in x, y and z direction (scale). As highlighted, the feature detection component is heavily dependent upon image differencing operators which generally are not ideal for SAR image statistics (speckle is multiplicative). Here, we don't intend to change the basic feature detection scheme but explore the idea of image smoothing and introduce the concept of skipping features detected at the highest scale with an objective to improve SIFT feature detection for SAR images.

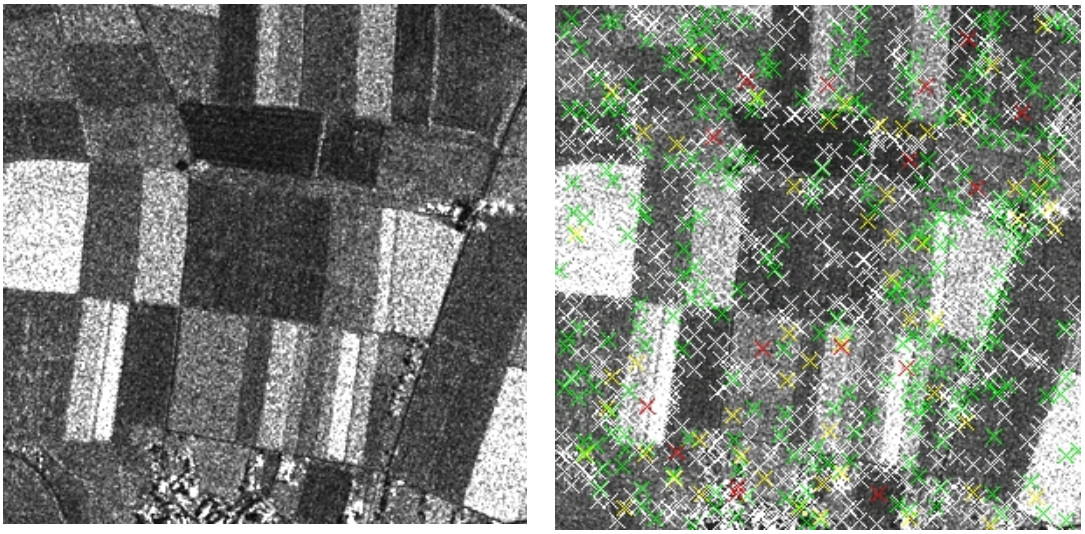
6.2.2 Feature Identification

Once stable features have been detected they are assigned robust descriptors which are scale and rotation invariant. The descriptors in the SIFT operator are assigned utilizing the local gradients computed at the characteristic scale of feature detection in the scale space pyramid. The feature descriptors depend upon local gradients estimated using simple adjacent pixel differencing and dominant gradient orientations of the feature neighbourhood are utilized to make the descriptor rotation invariant. More than one dominant orientation from the local gradient orientation histogram might be selected as the feature orientation and therefore numbers of descriptors might be greater than the number of features detected (recommended in Lowe, 2004). In this phase, the main step of estimating the local gradients is again not very well suited for SAR images and therefore we try to change the way local gradients are estimated. Later in this chapter we select

two conceptually different gradient operators (based upon image difference and ratio operator) specifically designed for SAR imagery.

6.2.3 Feature Matching

Matching of the feature descriptors based on ratio of the descriptor distance between two best matches of a reference feature (referred to as matching ratio in further text) seems to be a reasonable choice for both SAR and optical images. However, considering the properties of mutual information discussed in the previous chapters (see Section 3.4.4), we propose to combine MI and SIFT for improving the matching results obtained after using the standard SIFT operator matching (Section 6.9). In the following sections, we thoroughly evaluate the above proposed changes against the application of the standard SIFT processing chain for SAR images.



(a) (b)

Figure 6.1: Result of SIFT feature detection on a sample SAR image. (a) E-SAR imagery of size 300x300 pixels displayed at 75% of its original size. (b) The application of the SIFT operator produces 1520 features with around 74% features being detected at the first octave[13]. The figure depicts 1119 octave 0, 315 octave 1, 63 octave 2 and 23 octave 3 SIFT features marked by White, Green, Yellow and Red crosses respectively.

[13] For details see Appendix

6.3 FEATURE DETECTION

The application of the SIFT operator for SAR matching applications is not very straight forward primarily because of the multiplicative speckle influence inherent to any SAR image. Initiating the analysis, here we show the SIFT features detected on an E-SAR image (Experimental SAR: an airborne SAR-sensor developed and used by DLR (Schreiber et al., 1999)) of size 300 x 300 pixels (Figure 6.1). The application of the original SIFT operator on the image depicted in Figure 6.1a results into 1520 SIFT features being detected, depicted in Figure 6.1b. On analyzing the SIFT feature distribution in different octaves, it becomes evident that maximum features have been found in first octave and on their visual analysis it can be inferred that most of them are a result of the multiplicative speckle influence. Based on this initial observation we apply and test two modifications into the SIFT processing chain, described as follows:

Skipping of the first scale-space octave (SIFT-OCT): While analysing the SIFT operator performance for SAR image matching, it became apparent that only very few matches are found in the first octave (highest scale) of the scale space pyramid. Even worse, the matches found at the first octave were observed to have the highest false alarm ratio of all the octaves. Images at lower scale showed more robust features as the influence of speckle gets diminished lower down the octave pyramid. This observation suggests skipping the first octave of the scale space pyramid. The proposed change reduces the number of feature detections significantly, but interestingly it did not reduce the number of correct feature matches (demonstrated later). Another advantage is the reduced processing time as the image area reduces by four times in each octave.

Infinite symmetric exponential filter (SIFT+ISEF-OCT): To further reduce the speckle influence, the SAR images can be pre-processed using a smoothing filter before the computation of the scale space pyramid. The ISEF filter was shown to deliver good results for edge detection in SAR images (Fjørtoft et al., 1995). ISEF was proposed by Shen and Castan (1992) as an "optimal low-pass filter as a preparation for edge detection". Shen and Castan illustrated that an increase in Gaussian filter size is useful to reduce noise influence but the increased size has an adverse effect on edge localization precision. To overcome this problem, the ISEF filter with an

infinite window size and desired sharpness at the window centre was proposed. ISEF is mathematically expressed as

$$f(x) = \frac{p}{2}\exp^{-p.|x|} \tag{6.1}$$

Replacing p by $-\ln(b)$ the above equation can be written as

$$f(x) = ab^{|x|} \tag{6.2}$$

with $a = (-\ln(b))/2$ and $0 < b < 1$. For the two dimensional case, an efficient, recursive function can be used (Shen and Castan, 1992). The comparison between the Gaussian and the ISEF filter curves can be visualized in Figure 6.2.

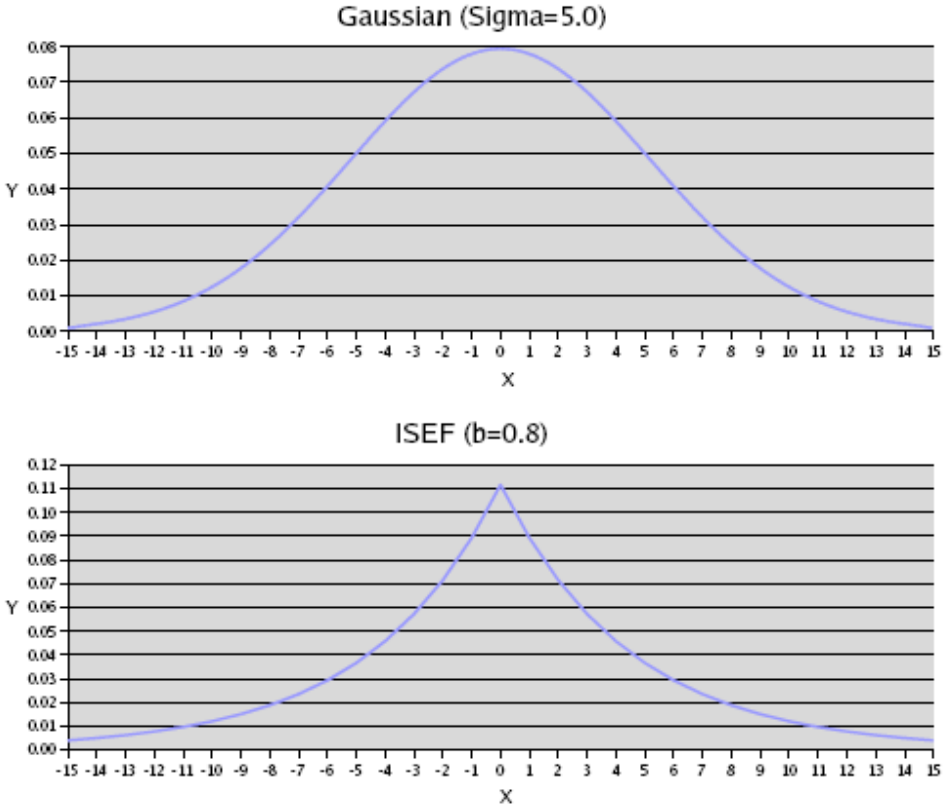


Figure 6.2: Comparison of Gaussian and ISEF function. The Gaussian function (top) smoothens at the centre and the ISEF filter (bottom) retains a sharp peak at the centre (Schwind, 2008).

6.4 FEATURE IDENTIFICATION

The two conceptually different gradient estimation techniques evaluated for improving the SIFT feature identification are the following:

Operator based upon image recursive differencing: Proposed for detecting step edges in noisy SAR images (Paillou, 1997), the utilized operator is derived from Deriche filter and is implemented recursively. The operator specifically designed for SAR images according to the edge detection criteria proposed by Canny (1983)

i. Insensibility to noise
ii. Good edge localization
iii. Unique response to one edge

was shown to be more effective than classical differential operator like Shen-Castan operator (Shen and Castan, 1986), first derivative of a Gaussian and the original Deriche operator (Deriche, 1987). For a detailed understanding of the operator, the interested readers are referred to the cited reference.

Operator based upon image ratio: Considering the fact that due to the multiplicative nature of speckle difference based edge detectors detect more false edges in areas of high reflectivity, we also test an edge detector with a Constant False Alarm Ratio (CFAR) (Touzi et al., 1988). For this comparison, we evaluate Ratio of Exponential Weighted Averages ROEWA operator proposed in Fjørtoft et al., (1995). This operator was found much efficient for SAR image edge detection in comparison to Ratio of Averages (Touzi et al., 1988), Ratio of Order Statistics (Brooks and Bovik, 1990) and ratio based version of Deriches optimal edge detector (Deriche, 1987).

The SIFT descriptor formation or feature identification as described in Appendix-1 is a function of a region of pixels selected for feature identification. It is clear from the nature of the task that the two major components to be scrutinized in this section are the magnitude of the

edge strength computed at a pixel followed by its assigned orientation. To recapitulate the magnitude Ma and orientations Or in image I are computed as following

$$Ma(i,j) = \sqrt{dx^2 + dy^2} \tag{6.3}$$

$$Or = a\tan 2(dy, dx) with - \pi \le a\tan 2(Y, X) \le \pi \tag{6.4}$$

where dx and dy represent the x and y gradient computed at a pixel location (i,j). Here on, we briefly outline the three different local gradient estimation techniques for a comparative evaluation.

Simple differencing (used in the original SIFT implementation)

The original SIFT implementation supports the simple first order derivatives calculated using the following equations

$$dx(i,j) = I(i+1,j) - I(i-1,j) \tag{6.5}$$

$$dy(i,j) = I(i,j+1) - I(i,j-1)) \tag{6.6}$$

Recursive differencing (Paillou, 1997)

This linear edge detector based on the Deriche filter uses the following recursive equations to obtain derivatives in x and y direction.

$$dx(i,j) = I_{xp}(i,j) + I_{xm}(i,j) \tag{6.7}$$

$$I_{xp}(i,j) = a_{0p} I_{x0}(i,j) + a_{1p} I_{xo}(i-1,j) - b_1 I_{xp}(i-1,j) - b_2 I_{xp}(i-2,j) \tag{6.8}$$

$$I_{xm}(i,j) = a_{1m} I_{x0}(i+1,j) +_{2m} I_{x0}(i+2,j) - b_1 I_{xm}(i+1,j) - b_2 I_{xm}(i+2,j) \tag{6.9}$$

$$I_{x0}(i,j) = a_1 \left(X_p(i,j) - X_m(i,j) \right) \tag{6.10}$$

$$X_p(i,j) = I(i,j-1) - b_1 X_p(i,j-1) - b_2 X_p(i,j-2) \tag{6.11}$$

$$X_m(i,j) = I(i,j+1) - b_1 X_m(i,j+1) - b_2 X_m(i,j+2) \tag{6.12}$$

In Equation 6.7-6.12, $i = 1...S_x, j = 1...S_y$, $a_{0p}, a_{1p}, a_{1m}, a_{2m,} b_2, b_1$ are operator parameters and S_x and S_y represent the image size. A similar recursive formulation is also utilized to obtain the gradient in y direction dy.

Ratio of exponential weighted averages (ROEWA) operator (Fjørtoft et al., 1995)

Placing a step edge between two homogeneous regions degraded by speckle, this operator intends to give an unbiased estimate of the ratio of the mean values of the two regions. Based on the ISEF filter (Figure 6.2b), which is realized as a cascade of two recursive filters, $h_1(x)$ (causal) and $h_2(x)$ (anticausal)

$$h(x) = c \cdot b^{|x|} \equiv h_1(x) * h_2(x) \tag{6.13}$$

where

$$h_1(x) = a \cdot b^x u(x) \qquad h_2(x) = a \cdot b^{-x} u(-x) \tag{6.14}$$

$c = a/(2-a), b = 1-a, 0 < a < 1$ and $u(x)$ is the Heaviside function. The first derivative of the smoothing filter $h(x)$ is given by

$$f(x) = \frac{d}{dx} h(x) \tag{6.15}$$

for ratio based implementation proposed in Fjørtoft et al., (1995)

$$f(x) = k\left(h_1(x) / h_2(x)\right) \qquad k = \tfrac{\ln(1-a)}{2-a} \tag{6.16}$$

For this case, horizontal and vertical components of the gradient of an image are given by:

$$\Delta_x\left[\left(h(x)h(y)\right) \otimes I(i,j)\right] = f(x) * \left(h(y) \bullet I(i,j)\right) \tag{6.17}$$

$$\Delta_y\left[\left(h(x)h(y)\right) \otimes I(i,j)\right] = f(y) * \left(h(x) \bullet I(i,j)\right) \tag{6.18}$$

where $\otimes$ denotes bidimensional convolution, $*$ denotes convolution in the horizontal direction and $\bullet$ represents convolution in the vertical direction. The discrete realizations of h_1 and h_2 are obtained using the z-transform of the discrete versions of equations (6.14)

$$H_1(z) = \frac{a}{1 - bz^{-1}} \qquad H_2(z) = \frac{a}{1 - bz} \tag{6.19}$$

For discrete recursive implementation (one dimensional), in terms of a spatial index n, this corresponds to the following simple recursions

$$s_1(n) = a\left(e_1(n) - s_1(n-1)\right) + s_1(n-1) \qquad n = 1....N \tag{6.20}$$

$$s_2(n) = a\left(e_2(n) - s_2(n+1)\right) + s_2(n+1) \qquad n = N....1 \tag{6.21}$$

In above, $e_1(n)$ and $e_2(n)$ are the inputs, and $s_1(n)$ and $s_2(n)$ are the outputs of h_1 and h_2 respectively. Adapting the edge detector to radar images, ratio $r(n)$ is defined as

$$r(n) = \min\left\{\frac{s_1(n)}{s_2(n)}, \frac{s_2(n)}{s_2(n)}\right\} \tag{6.22}$$

It is important to note that according to formulation represented in Equation 6.22, a stronger edge response would have value near to zero. Further, the Equations 6.20 and 6.21 are realized both for x and y directions to obtain the gradients dx and dy.

Table 6.1: Details of E-SAR Imagery used as dataset-1[14].

	E-SAR
Mode	Multi look image (4 looks)
Radar Frequency	9.6 GHz
Spatial Resolution	range: 2 m azimuth: 1.8 m
Pixel Spacing	1 m
Bits/Pixel	8
Incidence Angle	24.78°
Date of Acquisition	20/04/2004

For a comparative evaluation of the described gradient estimation schemes we perform the following tests

i. Run the three edge detectors on a sample E-SAR imagery to show the magnitude and orientation of edges produced by the three detectors.
ii. Run the same SIFT-OCT detection and the standard SIFT matching scheme on a sample dataset and compare the obtained results.

For evaluation, we have selected two scenes acquired by E-SAR sensor (Table 6.1) over Oberpfaffenhofen, Germany. Considering the small time difference and the use of the same sensor to acquire the images, a good registration result is anticipated. This dataset is also referred to as dataset-1 in further text. Sensor characteristics are tabulated in Table 6.1 and the two very similar images can be visualized in Figure 6.3.

[14] Two images acquired at 1 hr difference have been utilized as dataset- 1

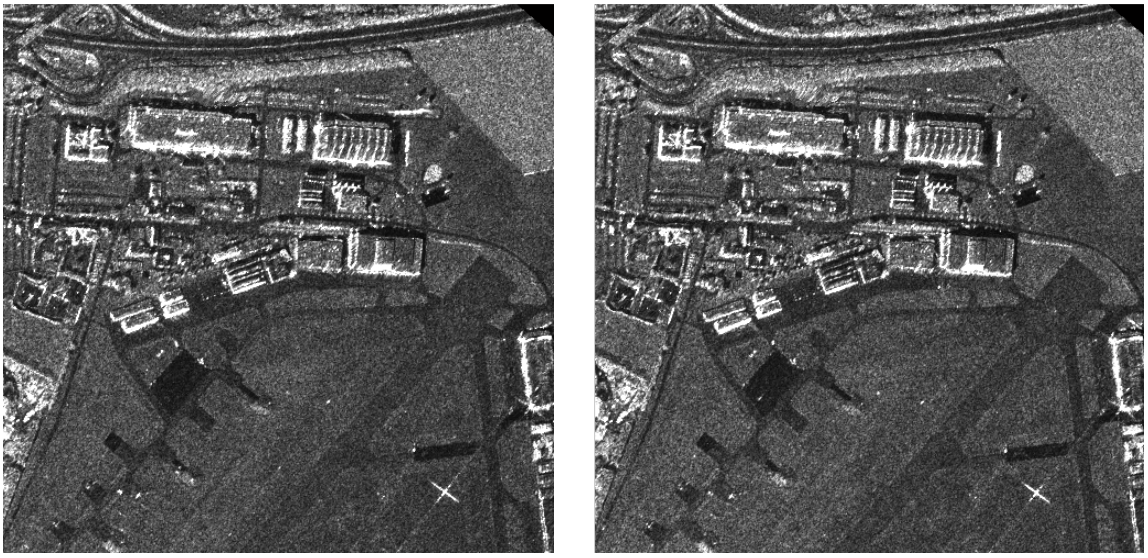
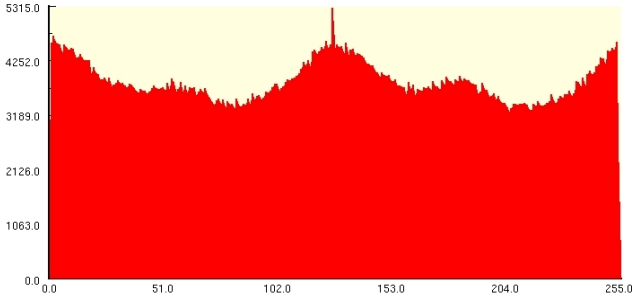


Figure 6.3: Two E-SAR images acquired over Oberpfaffenhofen, Germany at a time difference of 1hr.

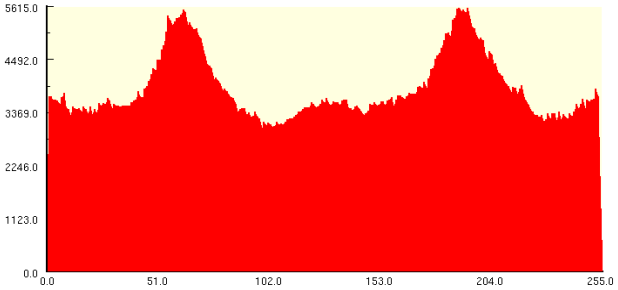
To analyse the gradient estimation schemes we show the individual gradient images and pixel orientation histogram obtained by the three schemes for the image shown in Figure 6.3a. The obtained results are provided for inspection in Figure 6.4. The magnitude images representing edge magnitude strengths are linearly mapped from their respective value range to the range $[0,255]$ (for display purposes). The edge orientations are again linearly mapped from the initial range $[-\pi,+\pi]$ to $[0,255]$ to plot a histogram of edge orientations of all the image pixels. The key observation to be made is that for difference based gradient estimation schemes the feature descriptors have orientations assigned in the complete viable range of $[-\pi,+\pi]$ but that is not the case for the ROEWA operator where all the image pixels obtain orientations from the range $[0,\frac{\pi}{2}]$. This behaviour is attributed to the Equation 6.22 where none of the x and y gradient has any sign associated which is used to decide an edge pixels orientation. This narrow range can influence the detector's performance especially for rotated images.

Magnitude image obtained by simple image differencing

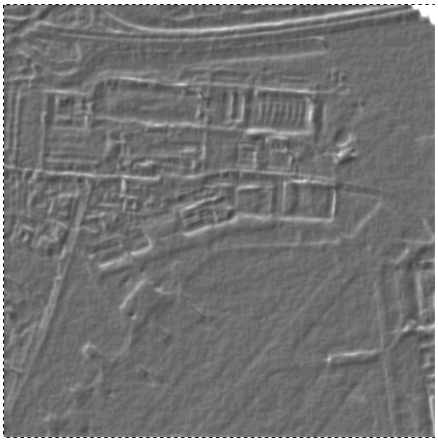


Edge orientation histogram depicting edges having assigned orientation in range $[-\pi, +\pi]$

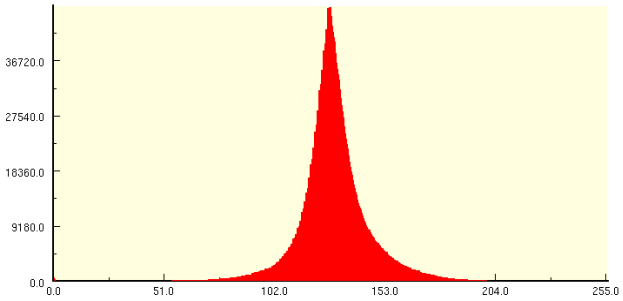
Magnitude image obtained by recursive image differencing



Edge orientation histogram depicting edges having assigned orientation in range $[-\pi, +\pi]$



Magnitude image obtained by ROEWA operator



Edge orientation histogram depicting edges having assigned orientation in range $\left[0, \frac{\pi}{2}\right]$

Figure 6.4: Edge magnitude and orientations obtained on applying different gradient estimation schemes on E-SAR imagery (Figure 6.3). The narrow range of orientation produced by the ROEWA operator might hamper its performance for images having a rotation difference. This is not observed for the gradient estimation techniques based on image differencing.

6.5 FEATURE MATCHING

To continue with the analysis we run the same SIFT-OCT detection, utilize the three different gradient estimators for matching images depicted in Figure 6.3 with the standard SIFT feature matching technique. For one matching scenario we also rotate the reference image by 90º to analyse how the gradient estimation schemes perform for rotated images. Further, we also propose another SIFT variant which does not have rotation invariance property but shows encouraging results for various tested datasets. SIFT descriptors rotation invariance can be easily given up by simply skipping the dominant orientation finding step in the SIFT processing chain (see Appendix) and this version of SIFT can be useful for remote sensing images which seldom have rotation differences.

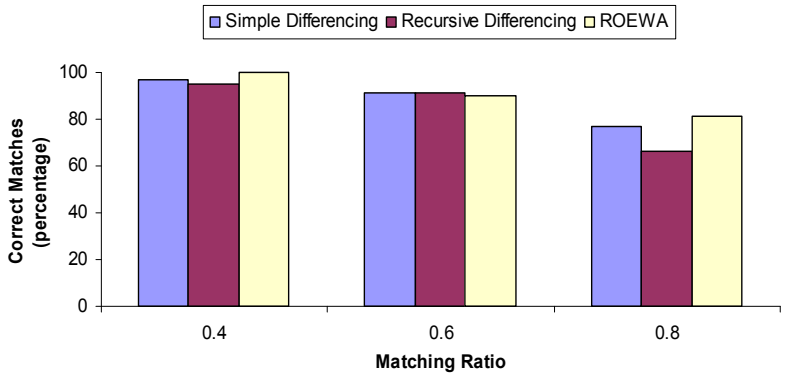
Table 6.2: Experimental results for difference and ratio based local gradient estimation techniques for E-SAR imagery in Figure 6.3.

	No Rotation			Reference image rotated by 90º			No Rotation		
Rotation Invariance	Yes			Yes			No		
Matching Ratio	0.4	0.6	0.8	0.4	0.6	0.8	0.4	0.6	0.8
Simple differencing	89	289	539	67	263	512	163	467	763
Recursive differencing	102	328	569	84	290	540	223	520	799
ROEWA operator	64	247	506	12	115	364	165	473	763

Due to the same detection scheme, we obtain the same number of detected features and similar matching accuracy levels in all the SIFT versions evaluated in this section, therefore these two factors have not been accounted for in the presented analysis. However, we compare the three techniques on basis of their matching accuracy in terms of percentage and number of correct matches. The observations with three different matching ratios (0.4, 0.6 and 0.8) are depicted in Figure 6.5. In terms of percentage accuracy of detection in all the three scenarios, it is seen that the originally proposed simple differencing performs almost at par with the two more conceptually sophisticated techniques (ROEWA operator shows a slightly improved performance).

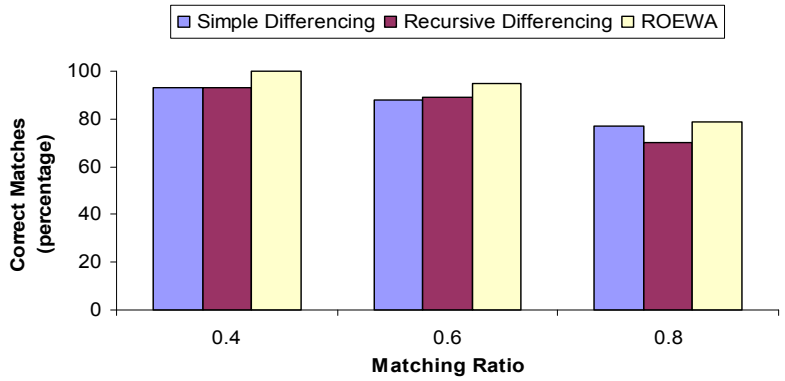
No rotation difference within the images

SIFT with rotation invariance



Reference image rotated by 90º

SIFT with rotation invariance



No rotation difference within the images

SIFT with no rotation invariance

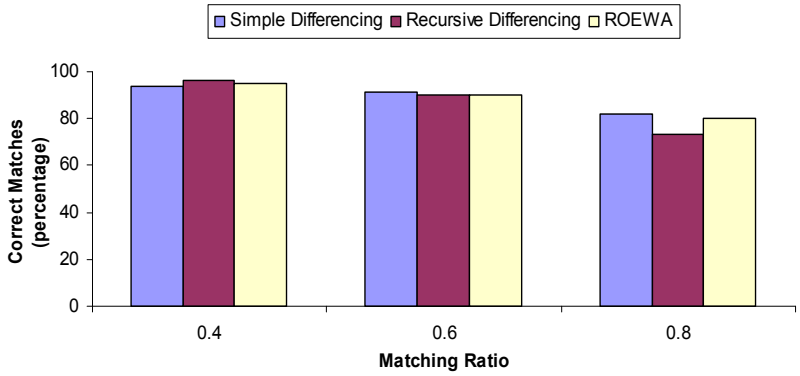
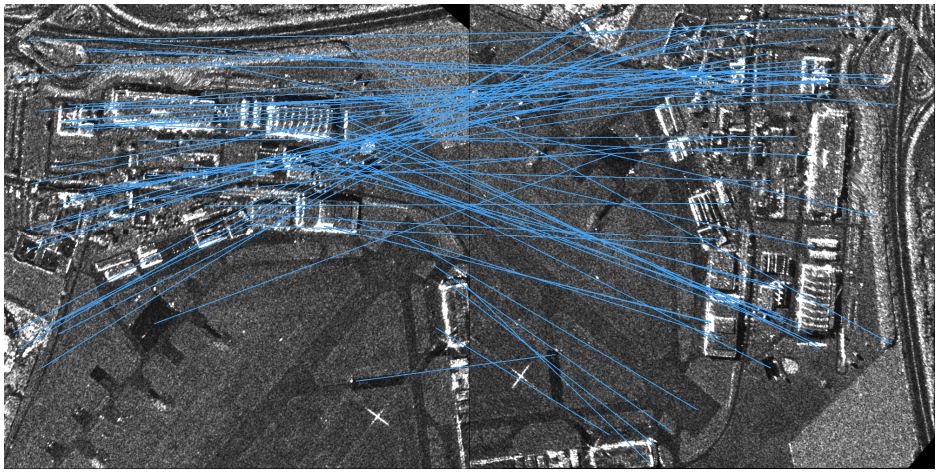
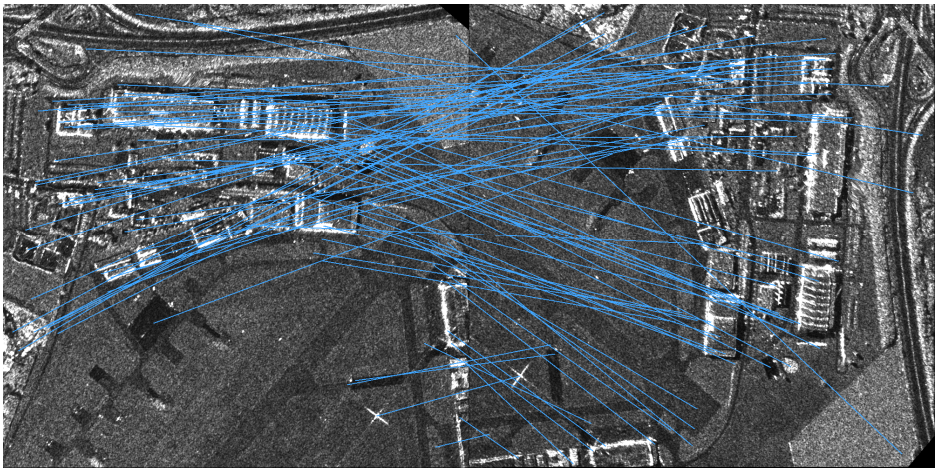


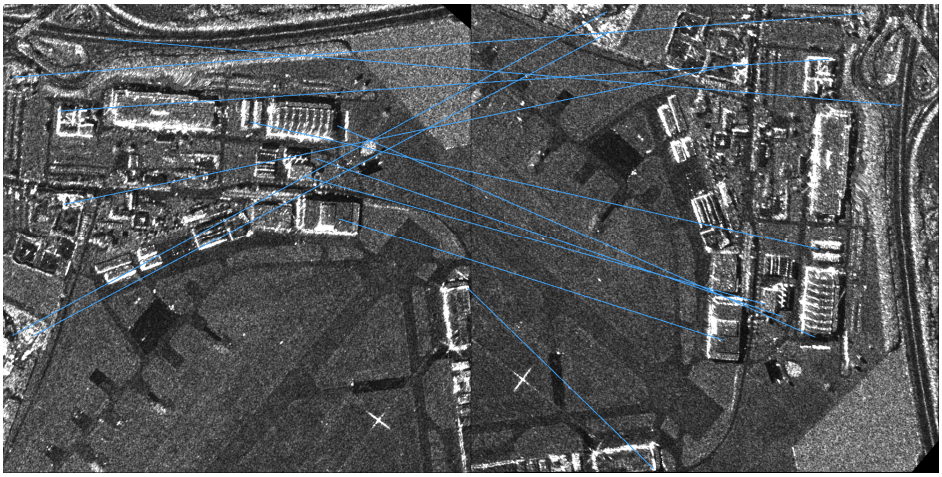
Figure 6.5: Comparison chart for difference and ratio based local gradient estimation techniques for E-SAR imagery in Figure 6.3. It is realized that the gradients estimated using simple image differencing perform at par with the two other much sophisticated techniques.



67 Matches found using simple differencing



84 matches found using recursive differencing



12 matches found using ROEWA operator

Figure 6.6: Matching results for E-SAR images having a 90º rotation difference. The number of matches for the ROEWA operator is far less as compared to the two other schemes after introducing a rotation difference of 90° within the images being matched.

Results in Table 6.2 indicate that the ROEWA based descriptor formation scheme looses maximum number of correct matches on rotating the reference image by 90º and this is due to the narrow range of feature orientations available (Figure 6.4). This problem is not encountered in the gradient estimation schemes based on image differencing. Similar performance trends have also been observed for several other datasets tested during the research phase with the ratio based scheme generally detecting considerably lower matches but with a slightly better accuracy percentage. Another key observation is the fact that the SIFT version with lost rotation invariance detects the maximum number of matches (for all the gradient estimating techniques). In general this has also been found true for various other datasets, suggesting the fact that the matching capability of the SIFT operator can be increased by giving away its rotation invariance property. Sample matches found between images having a rotation difference of 90º for all the three gradient estimation techniques are provided for visualization in Figure 6.6 (matching ratio 0.4). The results obtained indicate that gradients estimated by simple image differencing yield comparable performances to other sophisticated techniques, we further present a detailed SIFT operator evaluation (with rotation invariance) for diverse natured SAR images.

6.6 OPERATOR EVALUATION

6.6.1 Datasets

Results from three representative selected datasets (size 1000 x 1000 pixel) are being presented here (all the images analysed here are ground range amplitude images):

Dataset-1: Two E-SAR scenes acquired over Oberpfaffenhofen, Germany (Table 6.1).

Dataset-2: A dataset from multitemporal scenes acquired by Radarsat-1 (Canadian SAR satellite) over the city of Dresden, Germany, before and during the flooding of 2006. These scenes, showing a mostly urban land cover class were acquired at a time difference of almost six months (Table 6.3. A sample impact of the flood can be seen in the top left corner of Figure 6.10.

Dataset-3: Two scenes acquired using different sensors (Radarsat-1 and ERS-2, the European Remote-Sensing Satellite) at a time difference of 14/15 days, featuring a rural land cover class

with well defined features. The chosen scene shown in Figure 6.12 has prominent features (several lakes of the Lausitzer Seenkette near Senftenberg, Germany) to facilitate assessment of a feature based technique. To evaluate the operator performance for scenes with different aspect angles, the ERS-2 image was matched once with an ascending Radarsat-1 image (dataset-3a, Table 6.4) and once with a descending Radarsat-1 image (dataset-3b, Table 6.4).

Table 6.3: Details of Radarsat-1 imagery used as dataset-2.

	Radarsat-1 (ref)	**Radarsat-1 (inp)**
Mode	Standard Beam (mode 7)	Standard Beam (mode 7)
Radar Frequency	5.3 GHz	5.3 GHz
Pixel Spacing	12.5 m	12.5 m
Bits/Pixel	16	16
Incidence Angle	47°	47°
Date of Acquisition*	20/10/2005	06/04/2006
Orbit	Descending	Descending

Table 6.4: Details of Radarsat-1 and ERS-2 imagery used as dataset-3.

	ERS-2 (ref)	**Radarsat-1** (inp Dataset-3a)	**Radarsat-1** (inp Dataset-3b)
Mode	SAR Image Mode	Standard Beam (mode 7)	Standard Beam (mode 7)
Radar Frequency	5.3 GHz	5.3 GHz	5.3 GHz
Pixel Spacing	12.5 m	12.5 m	12.5 m
Bits/Pixel	16	16	16
Incidence Angle	22.97°	47°	47°
Date of Acquisition	20/04/2006	05/04/2006	26/04/2006
Orbit	Ascending	Ascending	Descending

6.6.2 Methodology

On the pre-registered images, the SIFT operator detection and the BBF (Best Bin First) matching technique has been critically analysed. We examine the original proposed SIFT operator and two of its minor variants (SIFT-OCT, SIFT+ISEF-OCT). The evaluation is done using the following criteria (Schwind et al., in print):

Number of detections: Individual feature detections made by the three detector variants have been recorded.

Overlap error (OE) and repeatability: To evaluate the detectors ability to detect same features in images with various acquisition differences (repeatability) the concept of overlap error between operator detected regions has been used. The SIFT operator detects regions centred at a feature point which is rotation and scale invariant. The scale at which the feature gets detected in the DoG pyramid is termed as the features characteristic scale (CS) (Lindeberg, 1998) and it determines the region size that is used for SIFT descriptor calculations. Regions around a detected feature might be circular or elliptical depending upon the nature of the feature detector (Mikolajczyk and Schmid, 2005). For scale and rotation invariant detectors like SIFT, Harris-Laplace (Mikolajczyk and Schmid, 2001), Hessian-Laplace (Lowe 2004), SURF (Bay et al., 2008) the region shape is circular. For affine invariant region detectors like Harris-Affine (Mikolajczyk and Schmid, 2004), Hessian-Affine (Mikolajczyk and Schmid, 2005) the affine region is elliptic and is determined using a shape adaptation process (Lindeberg and Garding 1997, Mikolajczyk et al., 2005). For the presented analysis we use circular detection regions where the radius of the circular region around the detected feature depends upon the scale of detection. A sample illustration of detected SIFT features at different characteristic scales along with their circular regions is provided in Figure 6.7 (For visualization ease, features detected at the first octave have not been depicted). It can be observed that the features detected at different scales provide circular regions of different sizes. The radius of the circular regions surrounding the feature point increase as we move down in the created scale space. The overlap error between these detected regions in different images has been employed as a criteria to evaluate the SIFT operator performance. Two regions are defined to correspond if the OE of these two regions is smaller than a set threshold ε (a threshold of 0.4 has been used) (Mikolajczyk et al., 2005)

$$1-\frac{R_{\mu_a}\cap R_{\left(H^T{}_{\mu_b}H\right)}}{R_{\mu_a}\cup R_{\left(H^T{}_{\mu_b}H\right)}}<\varepsilon \tag{6.23}$$

where R_μ represents the region and H is the homography relating the two images (a and b). $R_{\mu_a}\cup R_{\left(H^T{}_{\mu_b}H\right)}$ represents the union of the regions, and $R_{\mu_a}\cap R_{\left(H^T{}_{\mu_b}H\right)}$ is their intersection. The presented analysis has been performed on pre registered ground range images therefore the homography matrix H becomes an identity matrix for all the test cases. To counter the different region sizes, normalization is recommended (Mikolajczyk et al., 2005) and therefore a scale factor is determined to normalize the regions in the reference image to a radius of 30 pixels. The

computed scale factor is applied both to the region detected in the reference image and the region detected in the input image which is mapped onto the reference image before computing the overlap error. An illustrative Matlab code for normalizing the regions and computing overlap errors between the regions can be downloaded from www.robots.ox.ac.uk/~km/. The repeatability score for a given pair of images is computed as the ratio between the number of region to region correspondences and the minimum number of points detected in the images. For this analysis, the overlapping regions have been classified into four classes according to the OE. The classes are defined by the OE ranges 0-10%, 10-20%, 20-30%, 30-40% and higher than 40%. As all regions are scaled to a radius of 30 pixels (Mikolajczyk et al., 2005), an OE of 0-10% usually means a distance of 0-3 pixels between the region centres. Here it needs to be explained that some overlapping regions with a distance smaller than three pixels can still have an OE bigger than 10%, provided they were detected at very different scales. For an ideal detector performance, all of the overlapping regions would have an OE of 0-10%. In practice, such a result is unlikely. A more realistic but still favourable result would be to achieve maximum repeatability in the 0-10% class and sharply decreasing repeatability's for regions with an OE higher than 10%.

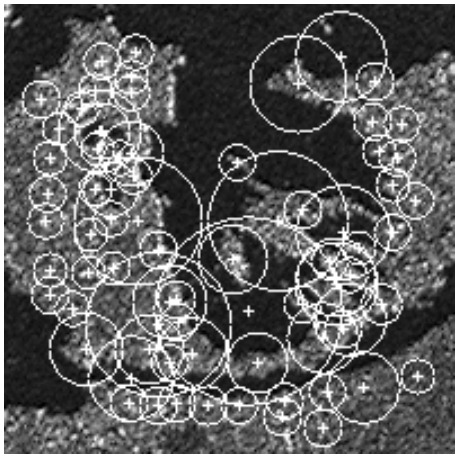


Figure 6.7: SIFT features displayed with their characteristic scale of detection in a SAR image. Radius of the circle centred on a detected feature is directly proportional to the feature detection scale. Here the features from the first octave have not been displayed to avoid cluttering.

Consistency analysis: To verify and evaluate the operator performance we compute polynomial transformation parameters (see Section 3.4.1) between the two feature sets detected in already

registered images. The average residual of the features obtained from the polynomial transformation has been used as an indicator of detector and the matching scheme performance.

Turn around time (TAT): Defined as the execution time needed by the detection and matching scheme to complete the task. All the experiments illustrated in this section have been performed on an Intel P4 Xeon (3 GHz, 1 MB cache) with 2 GB RAM.

The following section discusses in detail the experimental observations made from the three datasets using different SIFT operator variants. Here results obtained may not represent the optimum operator performance for a particular dataset but for a consistent and fair comparison, all the tests have been conducted with the same operator parameter settings.

6.7 EXPERIMENTS AND OBSERVATIONS

6.7.1 Dataset-1

The first dataset consists of same sensor images acquired at one hour time difference (Table 6.1). Therefore, good SIFT operator performance can be expected, the matches found in this dataset using SIFT-OCT variant can be visualized in Figure 6.8.

Number of detections: As observed in Table 6.5, the original SIFT operator detected 28660 and 28835 features in the reference and input image respectively. This huge number is drastically reduced while using the other two minor variants of the operator.

Repeatability: While analysing the repeatability measure it has to be noted that a high number of feature detections may lead to a misleading high repeatability score (Figure 6.9a). It is very likely that a possible match is found for every feature if both images are cluttered with features (see Note on the Effect of Region Density in Mikolajczyk et al., 2005). This can be seen as the reason for the original SIFT operator to show nearly the same repeatability score in the 0-10% and the 10-20% classes. A sharper peak difference for the 0-10% and 10-20% classes for SIFT-

OCT and especially for SIFT+ISEF-OCT illustrates the enhanced SIFT operator performance after the introduced modifications.

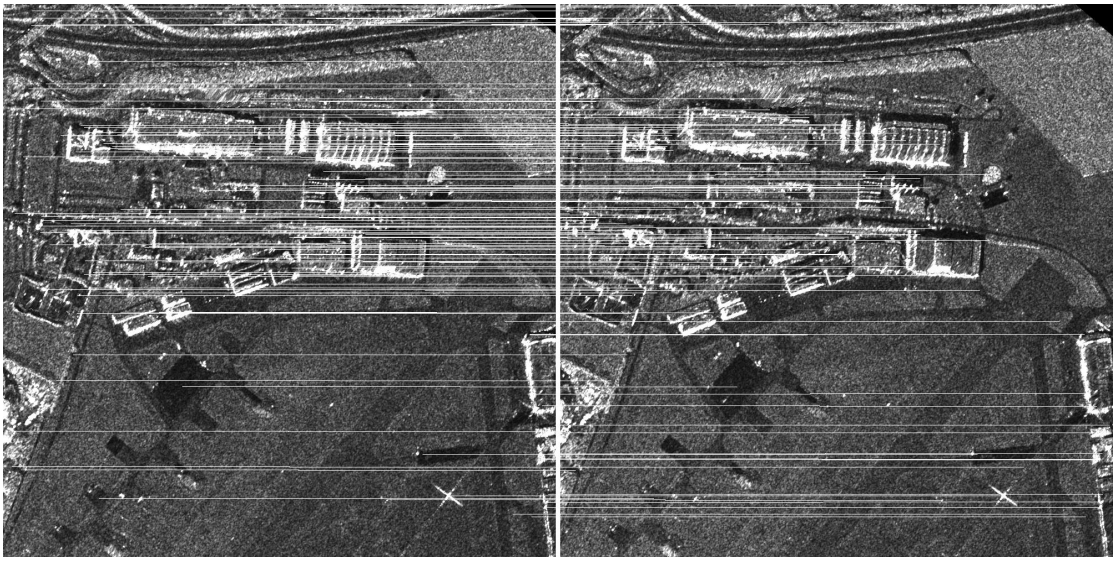


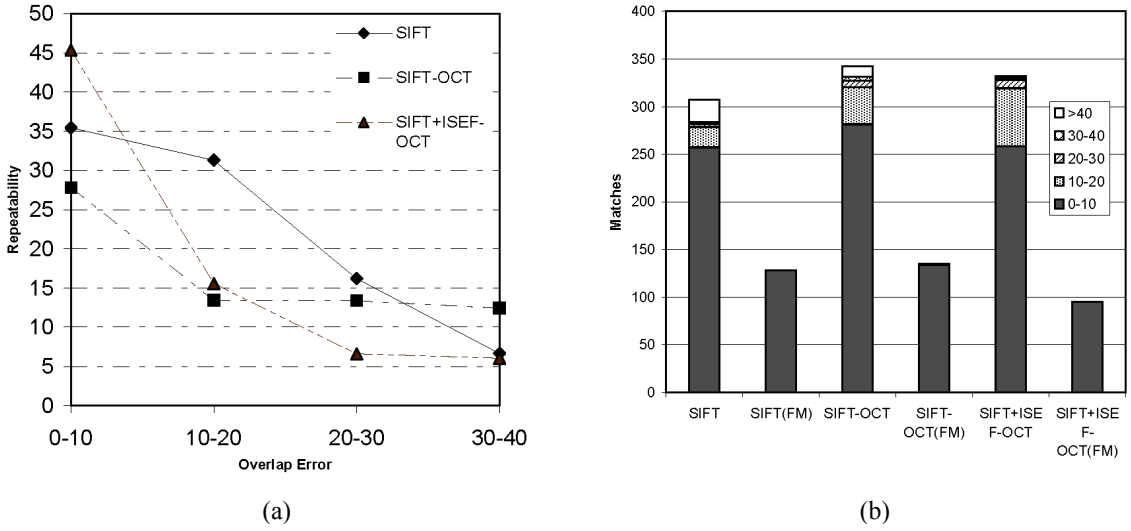
Figure 6.8: Matches found in dataset-1 after filtering them iteratively (SIFT-OCT scheme). The reference image is shown on the left, the input image on the right.

Table 6.5: Experimental results for dataset-1.

	Original SIFT	**SIFT-Oct**	**SIFT+ISEF-Oct**
TAT	201s	27s	17s
Nr. of features	Ref: 28660 Inp: 28835	Ref: 3942 Inp: 3938	Ref: 2048 Inp: 2153
Nr. of matches	307	343	332
Nr. of matches (FM)	128	134	95
Std. deviation	X: 74.72 Y: 47.25	X: 34.01 Y: 10.89	X: 1.95 Y: 1.30
Std. deviation (FM)	X: 0.26 Y: 0.30	X: 0.38 Y: 0.33	X: 0.38 Y: 0.26

Consistency: The significant number of outliers found by the BBF matching algorithm is responsible for the high standard deviations in X and Y returned by the polynomial image transformation (Table 6.5). To remove the outliers automatically, we use the same iterative approach utilized earlier (see Section 3.4.4). Using the residuals from an initial polynomial built using all the matched points, the point with maximum deviation from the polynomial is removed and a new polynomial is computed. This process which we call filtered matching (FM) is iterated until all returned residuals are smaller than twice the standard deviation of all the points.

Turn around time: The most obvious improvement of the modifications can be observed when taking into account the TAT of the detection and matching process. The SIFT operator without the first octave applied to the ISEF filtered image (SIFT+ISEF-OCT) is more than eleven times faster than the original SIFT operator (including the TAT for ISEF filtering). Skipping the first octave, which leads to a significant reduction in feature detections, can be mainly accredited for the speed up achieved.



(a) (b)

Figure 6.9: Repeatability score and number of matches for two same sensor images acquired at one hour difference (dataset-1). (a) The repeatability scores achieved for different overlap error classes. (b) Number of matches in different overlap classes (before and after the iterative filtering) for the three SIFT operator variants. SIFT feature repeatability capability increases significantly after the two proposed modifications in the detection scheme.

6.7.2 Dataset-2

The second dataset has same sensor images taken with a difference of six months. The matches found here using the original SIFT operator can be seen in Figure 6.10.

Number of detections: This parameter follows the same trend as we see a drastic reduction in number of detections with the variants of the original operator (Table 6.6).

Repeatability: As for the previous dataset, SIFT+ISEF-OCT achieves the sharpest peak difference between the 0-10% and 10-20% class (Figure 6.11a). Even though the total number of

overlapping regions is significantly lower for this dataset (Figure 6.7b), the repeatability score has remained consistently good showing that the temporal difference had little influence on detection quality.

Consistency: The initially high standard deviations returned by the polynomial transformation, are reduced to sub-pixel level if the outliers are removed iteratively (Table 6.6). Just as for the dataset-1, all matches with an OE larger than 10% were finally removed by the filtered matching (Figure 6.11b).

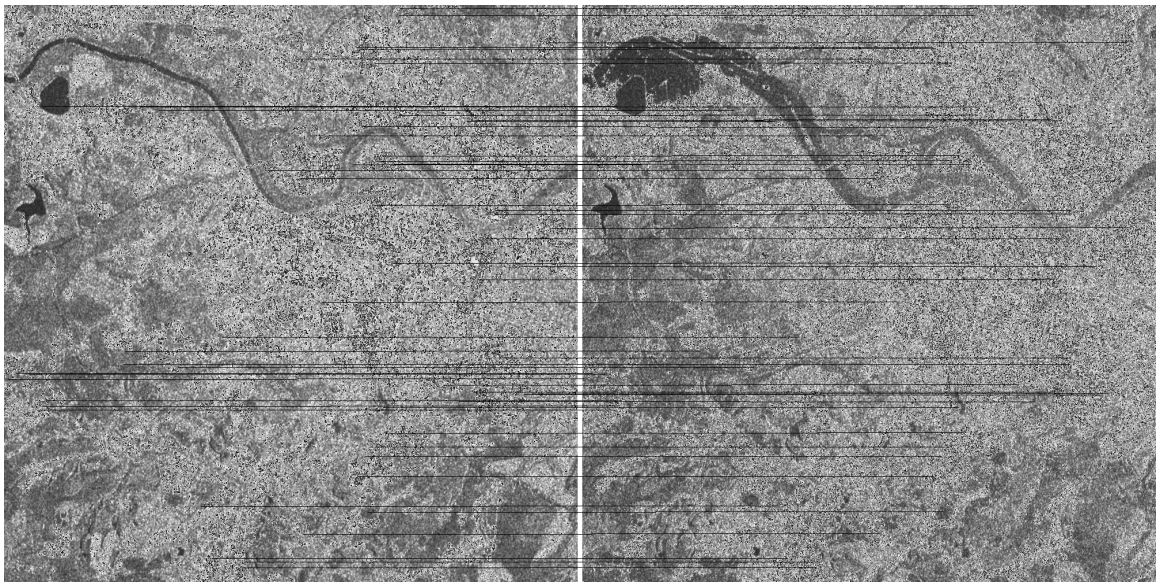
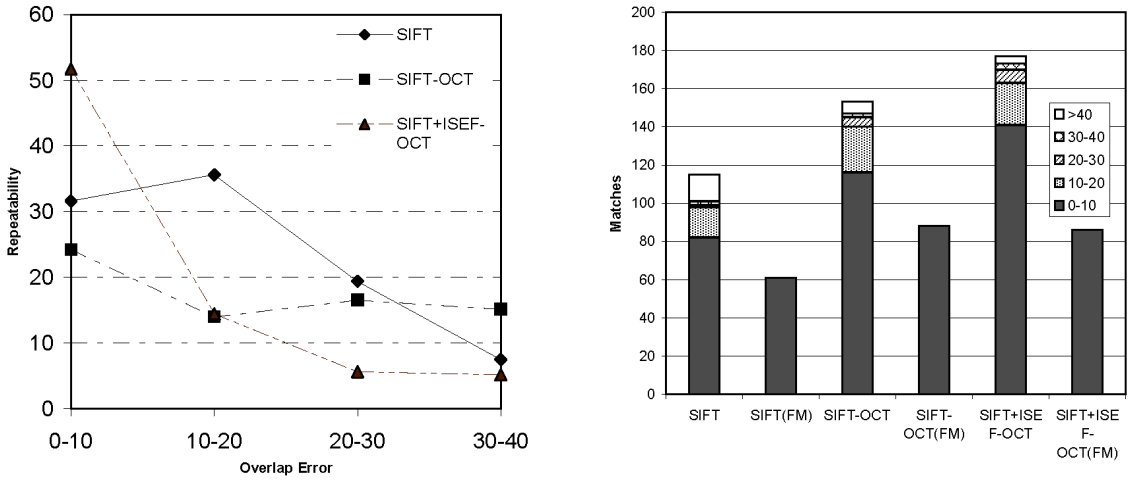


Figure 6.10: Matches found in dataset-2 after filtering them iteratively (SIFT scheme). Radarsat-1 reference image (left) acquired on 20/10/2005. Radardat-1 input image (right) acquired on 06/04/2006.

Table 6.6: Experimental results for dataset-2.

	Original SIFT	**SIFT-Oct**	**SIFT+ISEF-Oct**
TAT	180s	28s	19s
Nr. of features	Ref: 23681 Inp: 29282	Ref: 3881 Inp: 4381	Ref: 2382 Inp: 2622
Nr. of matches	115	153	177
Nr. of matches (FM)	61	88	86
Std. deviation	X: 124.05 Y: 150.06	X: 38.93 Y: 43.10	X: 39.47 Y: 54.89
Std. deviation (FM)	X: 0.64 Y: 0.54	X: 0.53 Y: 0.71	X: 0.44 Y: 0.55

Turn around time: SIFT+ISEF-OCT performs more than nine times faster than the original SIFT operator.



(a) (b

Figure 6.11: Repeatability score and number of matches for the multitemporal dataset-2. (a) The repeatability scores achieved for different overlap error classes. (b) Number of matches in different overlap classes (before and after the iterative filtering) for the three SIFT operator variants. As for dataset-1 in this case also the SIFT feature repeatability improves after introducing the changes in the feature detection scheme. It is worthwhile to note that the number of matches also increase in this case after the modifications.

6.7.3 Dataset-3

For the third dataset two separate tests were performed. First, an ERS-2 image acquired at an ascending orbit was matched with a Radarsat-1 image also obtained at an ascending orbit. Then, the same ERS-2 image was matched to a Radarsat-1 image of the same region acquired at a descending orbit. The matches obtained in dataset-3b using SIFT+ISEF-OCT can be seen in Figure 6.12.

Number of detections: The original SIFT operator by far has the most detection out of the three tested variants, while SIFT+ISEF-OCT finds the fewest features (Tables 6.7, 6.8).

Repeatability: The repeatability trend for both the scene pairs in dataset-3 does not represent ideal operator detection behaviour (Figures 6.13a and 6.14a). The fact that all the SIFT operator variants showed highest repeatability in the region of 10-20% overlap error indicates the influence of introduced image acquisition differences. In this case not

only the number of matches but the matching quality has also been affected. Considering dataset-3a, approximately one third of the matches returned (all the operator variants) have an OE bigger than 10% (Figure 6.13b).

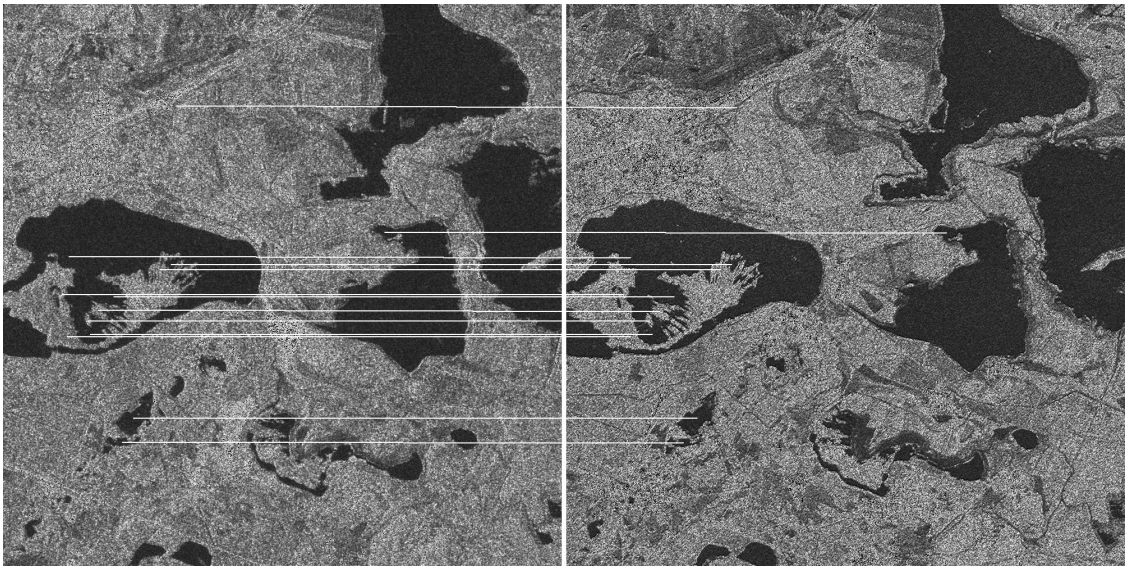


Figure 6.12: Matches found in dataset-3 after filtering them iteratively (SIFT+ISEF-OCT scheme). The reference image ERS-2 is shown on the left and the input Radarsat-1 image is on the right.

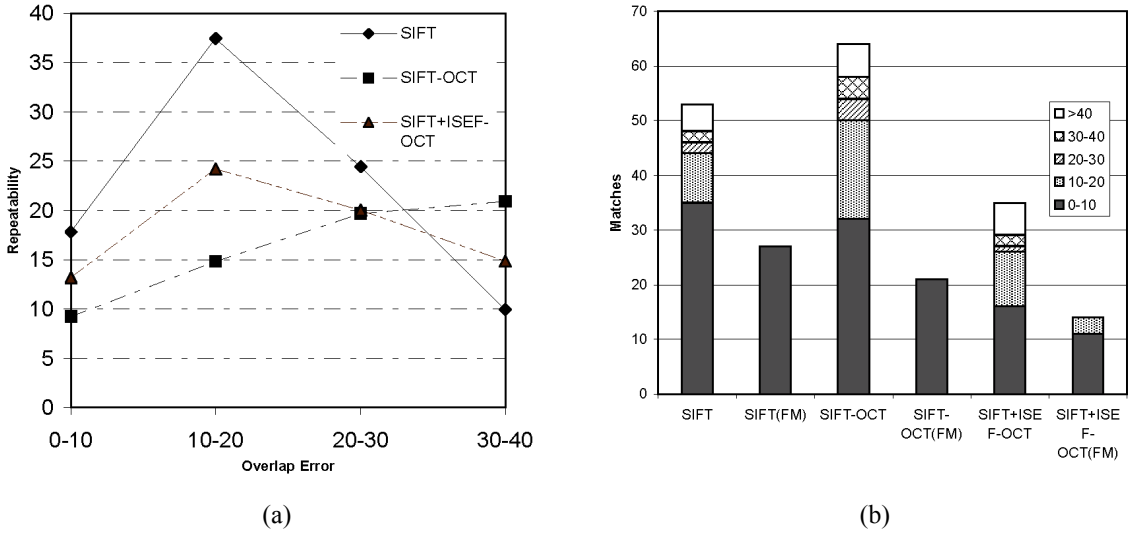
Table 6.7: Experimental results for dataset-3a.

	Original SIFT	**SIFT-Oct**	**SIFT+ISEF-Oct**
TAT	167s	25s	21s
Nr. of features	Ref: 29876 Inp: 24931	Ref: 4061 Inp: 3426	Ref: 3725 Inp: 3247
Nr. of matches	53	64	35
Nr. Of matches (FM)	27	21	14
Std. deviation	X: 125.90 Y: 84.23	X: 18.93 Y: 90.46	X: 125.41 Y: 111.65
Std. deviation (FM)	X: 0.53 Y:0.50	X: 0.31 Y: 0.58	X: 0.70 Y: 0.33

Table 6.8: Experimental results for dataset-3b.

	Original SIFT	**SIFT-Oct**	**SIFT+ISEF-Oct**
TAT	169s	26s	21s
Nr. of features	Ref: 29876 Inp: 24931	Ref: 4061 Inp: 3204	Ref: 3725 Inp: 3227
Nr. Of matches	24	22	20
Nr. Of matches (FM)	12	10	13
Std. deviation	X: 262.49 Y: 184.80	X: 139.81 Y: 87.19	X: 111.81 Y: 107.26
Std. deviation (FM)	X: 2.07 Y: 1.16	X: 1.84 Y: 0.54	X: 1.47 Y: 0.98

Still, iterative filtering in this dataset led to favourable results as all the finally retained matches were in the overlap error of 0-10% (expect for SIFT+ISEF-OCT, 3 matches left with OE of 10-20%). The match performance further deteriorated with added complexity of ascending/descending orbits (dataset-3b) as here matches with OE of 0-10% made up less then half of all the matches found (Figure 6.14b) and iterative filtering could not bring all the matches to the desired overlap error.

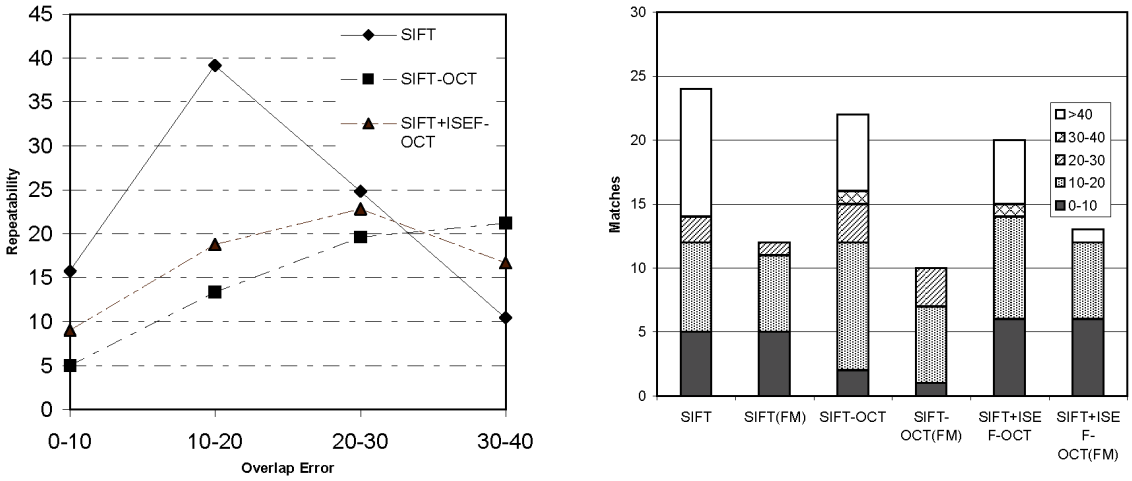


(a) (b)

Figure 6.13: Repeatability score and number of matches for the multisensor dataset-3a (both scenes acquired with ascending orbits). (a) The repeatability scores achieved for different overlap error classes. (b) Number of matches in different overlap classes (before and after the iterative filtering) for the three SIFT operator variants. The feature repeatability for all variants shows peak in the overlap category of 10-20% (unlike the last two datasets) mainly due to very different sensor characteristics. Still enough number of matches with acceptable consistency found for image registration applications.

Consistency: As done with the previous datasets, the outliers have to be filtered out to achieve higher consistency (Tables 6.6, 6.7). The final number of matches for all of the three variants has been reduced significantly. Even though sub-pixel consistency could be achieved for dataset-3a with a descent number of matches, this was not the case for dataset-3b (match consistency of around 2 pixels). The lack of sub pixel consistency in the matches can be related to deteriorated feature localization (observed in repeatability analysis, Figure 6.14b) and the significant drop of number of matches might be attributed to changes in feature descriptors which have been

significantly affected by the further introduced acquisition difference of ascending/descending orbits.



(a) (b)

Figure 6.14: Repeatability score and number of matches for the multisensor dataset-3b (images acquired with ascending/descending orbits). (a) The repeatability scores achieved for different overlap error classes. (b) Number of matches in different overlap classes (before and after the iterative filtering) for the three SIFT operator variants. Again, the feature repeatability for all variants shows peak in the overlap category of 10-20% and number of matches are further reduced (compared to dataset-3a), asserting the influence of different orbits of Radarsat and ERS utilized for image acquisition.

Turn around time: SIFT+ISEF-OCT again achieved a significant speed-up over the original SIFT operator, performing almost eight times faster.

6.8 ASSESMENT OF EXPERIMENTS

6.8.1 Feature Detection

As observed in Tables 6.5-6.8, the SIFT operator detects an undesirable high number of features for all the four test cases. This can primarily be attributed to high amount of speckle normally present in SAR images. To counter the speckle influence, the ISEF filter is selected for image smoothing and reducing the number of features detected. As speckle influence is likely to dominate at the highest image scales, the features detected at first octave are also not considered for matching. As a consequence, the numbers of features detected are reduced considerably for

all the datasets. Further, the observation that the SIFT-OCT finds a higher number of correct matches (almost same consistency) in a smaller number of detected features (except for dataset-3 where the number of matches is slightly lower with better consistency) highlights the adverse speckle influence on feature detection and matching.

6.8.2 Feature Repeatability

The repeatability measure provides statistical insights into the SIFT operator capability to detect similar features in images with various differences. For a very good detector performance, the repeatability of key point detection should be the maximum in the overlap region of 0-10% and the repeatability score in other overlap error regions should decrease successively and finally should become equal to zero. For dataset-1 and 2 (Figures 6.9a and 6.11a) the two SIFT operator variants (SIFT-OCT and SIFT+ISEF-OCT) performs nicely, yielding maximum repeatability scores in the overlap error category 0-10% and decreasing thereafter. The same trend is observed for the original SIFT operator for dataset-1 but not for dataset-2. Where, the repeatability score for the overlap category 10-20% increased a bit before following the decreasing trend. For dataset-3 (Figures 6.13a, 6.14a) all the tested operator variants reported maximum repeatability score for overlap category 10-20%, deviating from the favourable performance shown in the two previous datasets decreasing thereafter. In principle, the SIFT features are rotation and scale invariant and robust enough for affine transformations but still their localization and descriptors get significantly affected by their local neighbourhood. Specifically for SAR images, a potential SIFT feature's neighbourhood, eventually deciding its position (sub pixel localization) and identification (descriptor) is prone to changes with changing sensors and/or acquisition conditions (incidence, aspect angle, polarization etc). The influence of the changed neighbourhood (dependent upon acquisition conditions) has definitely influenced feature localization for both the scene pairs in dataset-3 and is observed through the deteriorated repeatability performance of the SIFT operator (Figure 6.13a, 6.14a). Feature descriptors have also suffered in both the scene pairs and this is reflected by lower number of matches obtained (Figure 6.13b, 6.14b) when compared to the number of features matched in the previous datasets. Comparing the results for the two scene pairs of dataset-3 in isolation further supports the assumed hypothesis. Here for a relatively flat terrain, the repeatability, number and quality of

features matched get reduced significantly for dataset-3b which can be seen as a result of changed feature neighbourhoods due to the effect of different sensor orbiting directions introduced on top of different sensor and incidence angle of acquisition (dataset-3a).The magnitude of possible changes will further depend upon the nature of the area being imaged. For example, different incidence and aspect angle dependencies are expected to have a stronger impact on potential SIFT features in urban (especially for metric resolution images) and mountainous regions as compared to the features in plain areas.

6.8.3 Feature Matching

Considering only the feature repeatability measure would not complete the SIFT operator analysis for its applicability to SAR image registration. As the feature SIFT descriptors have been matched using the BBF matching technique, we also evaluated the final quality of the matched features. The BBF matching technique has been found not to be 100% reliable especially for matching of remote sensing SAR images. As can be seen in all the three datasets tested and analysed (Tables 6.5-6.8), the BBF technique for all the three variants matched features with different overlap errors. Further, the polynomial built using the initially returned matched features in all the cases lead to unacceptable average residuals. The matches (without filtering) for datasets -1, -2 and -3b from one of the operator variants can be seen in Figures 6.8, 6.10 and 6.12 respectively. The results returned by the matching technique become useful after an iterative rejection of outliers. It can be seen from the results obtained for dataset-1 and dataset-2 that after the iterative elimination, all the matched points from the three SIFT operator variants get restricted to the overlap category 0-10% (Figures 6.4b and 6.6b). Similar performance was also observed for dataset-3a where a small anomaly is detected for the SIFT+ISEF-OCT variant (Figure 6.13b).

The matching results of the multisensor dataset acquired under different orbiting directions (dataset-3b) shows the maximum deviation from the results obtained for all other test cases. Because of the introduced aspect angle dependency, deteriorated SIFT operator performance is reflected in the distribution of matches in overlap error categories (Figure 6.14b). It can still be argued that the tested operator settings might not be optimal for the dataset under consideration

but a better and a more reasonable choice for such cases would be to employ intensity based metric like mutual information to refine the matched points using chip matching techniques (see Section 3.3.4).

6.8.4 Process Turn around Time

Here, process turn around time incorporates the detection and matching process execution times. From Tables 6.4-6.7, it is observed that image smoothing and skipping the first octave provides a significant speedup advantage over the original SIFT operator. The reason for the attained speed up lies in the fact that fewer features are detected and finally matched.

6.8.5 Applicability Recommendations

On basis of the presented experimental results and their critical analysis based on feature detection, repeatability, matching and process turn around time the following conclusions can be drawn:

i. The SIFT operator with the BBF matching technique has shown potential to be a stable method for point feature based registration of SAR images. Subpixel consistency was achieved by one or more SIFT operator variants for all the test cases, except dataset-3b (different sensor, incidence angle and sensor orbit direction).

ii. On the basis of the tested datasets, it can be inferred that the SIFT operator detection and the BBF perform best for same sensor images with decreasing performance as the acquisition differences between the images increase.

iii. The two SIFT operator variants have also shown a significant speed up with similar consistency levels compared to the original SIFT operator.

iv. On basis of the SIFT operator performance and the analysis of MI provided in Chapter 3, a hybrid multimodal SAR image matching scheme combining the benefits of both is

developed and discussed in the subsequent section with a challenging image matching scenario.

6.9 A HYBRID CHAIN FOR SAR IMAGE MATCHING

The concept of combining intensity based technique like MI and feature based technique like the SIFT can be utilized for complex image matching/registration scenarios. We present an argument that all the correct SIFT matches are not reported in the obtained results, mainly due to matching ratio threshold parameters. This is confirmed by the fact that the SIFT operator show very high feature repeatability in terms of absolute number of detection, especially in the overlap category 0-10% for different registered datasets in the previous section. But very few of those are actually reported as successful matches after the matching phase. In general, detection of all possible correct matches might not be very relevant always as the SIFT processing chain with suitable threshold parameters can detect enough number of correct matches for accurate registration (as shown in the previous section). However, the problem might arise for very complicated image matching scenarios like multitemporal registration (see Section 1.1) where the SIFT chain might not produce enough number of matches to meet the desired objective mainly due to drastic changes in scene conditions (It is already highlighted through dataset-3 that SIFT feature detection and matching results deteriorate for images with complex acquisition differences). For such scenarios, we propose to utilize MI and assist the SIFT matching scheme to find out all possible correct matches within images. MI can quickly estimate rough shift parameters from compressed images and this can be utilized to increase the number of matches in SAR imagery using the SIFT processing chain. This concept is depicted in this section through a hybrid SAR image matching chain (Figure 6.15) tested on a complex multitemporal/multisensor dataset.

6.9.1 Dataset

The chief characteristics of this selected dataset can be summarized as:

i. Different sensor nature (airborne and spaceborne).
ii. Time difference of 4 years.
iii. Different incidence angle.

iv. High resolution imagery acquired over semi urban land cover.

The mentioned image acquisition differences present a complicated scenario for image matching applications. The influence of different sensor geometries and incidence angles is expected to bring strong aspect dependency on the appearance of urban settlements in the scene.

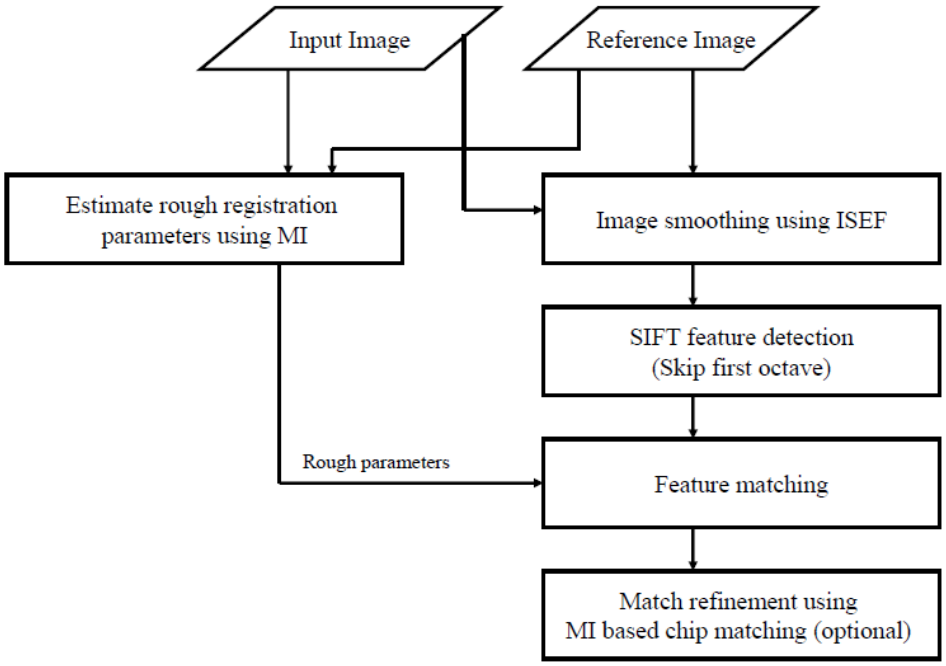


Figure 6.15: A hybrid processing chain combining mutual information and scale invariant feature transform for SAR image matching applications (Suri et al., 2009a).

Table 6.9: Details of the E-SAR and the TerraSAR-X Imagery.

	E-SAR	**TerraSAR-X**
Mode	Multi-look image (4 looks)	High resolution spot light
Radar Frequency	9.6 GHz	9.6 GHz
Pixel Spacing	1 m	1 m
Bits/Pixel	8 bit	16 bit
Incidence Angle	24.78°	35.14°
Data of Acquisition	20-April-04	20-May-2008
Orbit	-	Ascending
Image Size	1000 x 1000	1000 x 1000
Product type	-	GEC

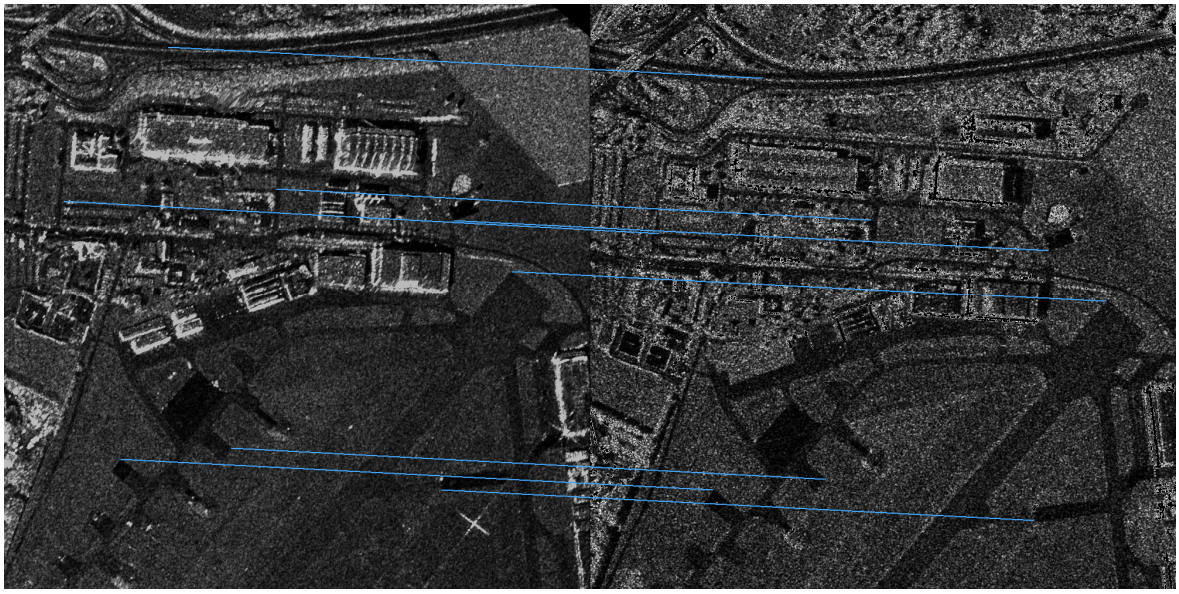


Figure 6.16: The images from (left) airborne E-SAR (right) TerraSAR-X acquired over Oberpfaffenhofen near Munich in Germany. The images have approximately 4 years of acquisition difference. The end points of the blue lines represent the eight[15] tie points utilized for registration accuracy assessment.

The images have been taken over Oberpfaffenhofen (near Munich) by E-SAR sensor and the latest German high resolution satellite, TerraSAR-X. The different sensor nature and incidence angles is expected to have an influence on the features in the urban areas and thus this selection of dataset offers an interesting evaluation of the SIFT operator capability to match and detect features in SAR images with complex acquisition differences. The details of the imagery selected are tabulated in Table 6.9 and the changes in the sub urban area in form of new constructions and deconstructed buildings in 4 years time span can be observed in Figure 6.16 (images not registered). As the reference TerraSAR-X product is not an EEC product, we see the scene deviating from the ground reference by approximately 12 m in the x direction and 3 m in the y direction. To evaluate the registration performance of MI, SIFT and their combination we have manually marked eight points in the two images, using these points for building an affine transformation we obtain a standard deviation of 1.14 m in x direction and 0.71 m in the y direction. Shifts around these eight points are used for accuracy assessment of the registered images obtained using different methodologies.

[15] Lines from two tie points coincide that is the reason that only seven prominent lines are visiible in the image

6.9.2 Results

Mutual information: We start our analysis showing results from the intensity based registration process. For a rough approximation, the transformation model for the MI based process has been limited to translation in x and y directions. The optimization process has been initiated with the seed (-10, -45) and for mutual information computations, a joint histogram of bin size 64 has been estimated using the GPVE technique (Suri and Reinartz, 2008). Table 6.10 tabulates the mutual information performance for the selected dataset as described in the earlier chapters for SAR and the optical case for multimodal SAR image registration also MI could estimate very similar registration parameters in far less execution time from down sampled images of 4 and 2 m pixel spacing respectively.

Table 6.10: MI based registration of TerraSAR-X and E-SAR imagery.

Pixel Spacing	TSAR-X to E-SAR[16]	E-SAR to TSAR-X	RC	TAT[17] (sec)
4 m	(-13.76, -52.60)	(13.76, 52.56)	0.00	70
2 m	(-13.78, -52.58)	(13.76, 52.56)	0.12	150
1 m	(-14.80, -52.61)	(15.01, 52.90)	0.36	440

In Table 6.10, we also provide the metric consistency measure (see Section 2.4.3) normally utilized for intensity based registration processes in absence of ground truth measurements. Normally values near to zero of the Registration Consistency measure (RC) represents a good metric performance. Important point to be kept in mind is that the value of the consistency measure might not be directly related to the ground level accuracy of registration achieved.

SIFT operator: We analyse here the SIFT operator matching scheme performance both with and without using the rough registration parameters in the SIFT processing chain. The matches obtained at matching thresholds of 0.6, 0.8 and 1.0 (see Section 3.3) have been reported for the analysed dataset in Figure 6.11. As explained earlier, the matching threshold is calculated as the ratio between Euclidean descriptor distance of the second-closest and the closest match of a feature. At matching ratio 1.0 we utilize the rough registration parameters to compute an

[16] Transformation parameters represent translation in x and y direction (in pixels)
[17] On an Intel P4 Xeon (3 GHz, 1 MB cache) with 2 GB RAM

approximate match region for every match and filter out those matches where corresponding feature don't lie within a user defined window size (16 pixels). Finally, depending upon the number of matches or match consistency and application demands it can be further decided to refine each of the matches individually by using MI locally around the matched features (chip matching technique).

As shown for the Landsat-Radarsat registration scenario (see Section 3.3.4) a simple shift estimation within large remote sensing scenes[18] may not suffice for a precise registration, which can normally be achieved using an affine transformation. Taking the initial guess obtained from level 2 images we continue with the SIFT operator performance analysis. The effect of complex acquisition differences mentioned above is observed on the SIFT matching results tabulated in Table 6.11. The standard matching procedures with matching ratios of 0.6 and 0.8 produced far less matches as compared to the matches produced by matching ratio of 1.0 helped by the initial estimate provided by the intensity based process. Considering the nature of the datasets the 57 matches found by the SIFT operator are encouraging (both on ground and within urban establishments).

Table 6.11: SIFT matching scheme performance.

Matching Ratio	**0.6**	**0.8**	**1.0**	**1.0** "on ground"
Matches Found by SIFT	7	188	57	38
Filtered Matches (RANSAC)	3	10	57	12

The matched features for this dataset can be visualized in Figure 6.17 roughly classified into plain area features (green lines) and features matched within the urban establishments (blue lines). All the 57 SIFT features detected here, might be useful for various SAR image matching scenarios but for co-registration only on ground features (38 green colour lines in Figure 6.17) should ideally be considered for registration parameter estimation. In isolation, the "on ground" 38 features have a consistency of around 2 pixels in both x and y direction and Equation 6.23 represents the estimated affine transformation.

[18] The current dataset has 1 m pixel spacing images of size 1000x1000 pixels so is relatively much smaller ground area covered by images in Section 3.3.4

$$\begin{bmatrix} TSARX_X \\ TSARX_Y \\ 1 \end{bmatrix} = \begin{bmatrix} 1.00 & -0.45E-02 & 16.01 \\ -0.38E-02 & 1.00 & 51.33 \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} ESAR_X \\ ESAR_Y \\ 1 \end{bmatrix} \quad (6.23)$$

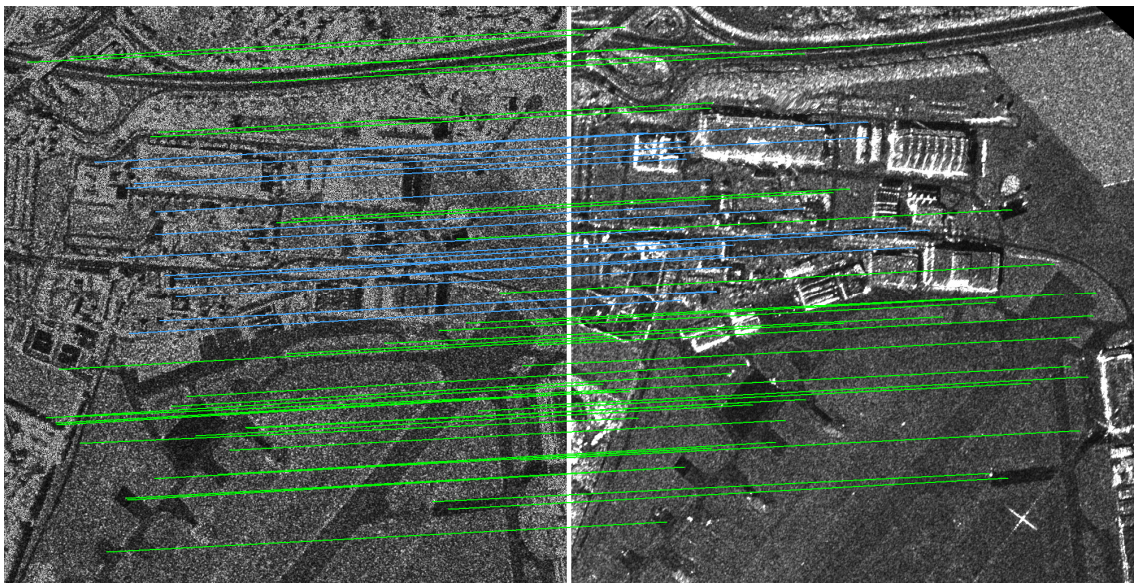


Figure 6.17: SIFT features matched between transformed TerraSAR-X (left) and E-SAR image (right). In total 57 SIFT features were found using a matching threshold of 1.0 with the rough registration parameters from intensity based techniques. Green lines represent features matched on plain ground and blue lines represent features matched within the urban establishments (Suri et al., 2009a).

SIFT and MI: For this test we could refine 19 conjugate features out of the 38 "on ground" features as a window of size 300 x 300 pixels needs to be demarcated centred on each feature. Thus, features lying near the image boundaries can not be considered for further refinement. Finally, we obtain 12 conjugate features (after outlier removal) with a standard deviation of 0.75 pixels in x and 0.54 pixels in y direction. The obtained affine parameters keeping TerraSAR-X image as the reference are shown in Equation 6.24.

$$\begin{bmatrix} TSARX_X \\ TSARX_Y \\ 1 \end{bmatrix} = \begin{bmatrix} 0.99 & -0.63E-02 & 19.13 \\ -0.5E-02 & 1.00 & 50.65 \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} ESAR_X \\ ESAR_Y \\ 1 \end{bmatrix} \quad (6.24)$$

Accuracy assessment: For a comparative evaluation of the results produced by the three techniques, we utilize the eight manually marked check points shown in Figure 6.16. The results

are provided using a scatter plot in Figure 6.18 and corresponding statistics are provided in Table 6.12. Here, we observe that the SIFT operator performance alone seems to be the worst amongst the three methods under consideration. Utilizing MI alone for this 1000x1000 pixel image having 1 m pixel spacing has given satisfactory results but it has been shown earlier that estimating simple translation shifts may not necessarily be sufficient for large remote sensing scenes.

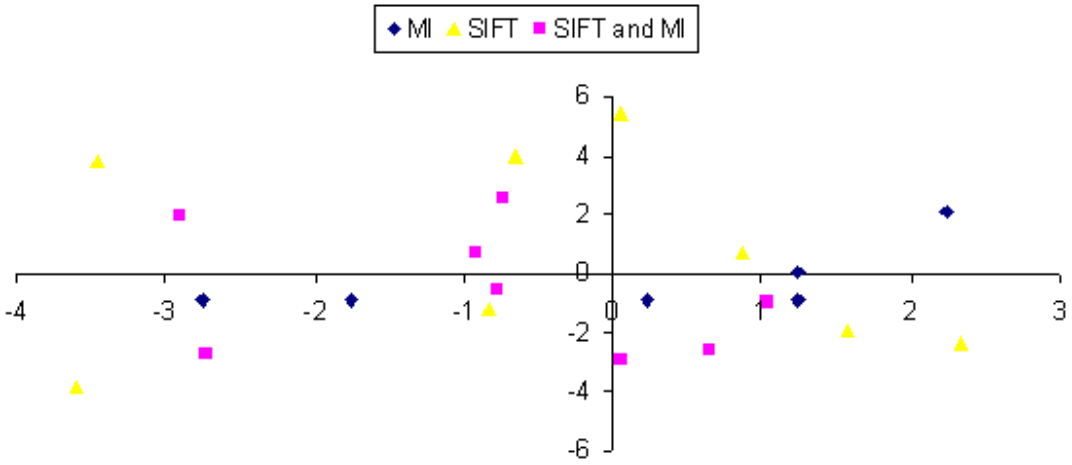


Figure 6.18: Scatter plot of difference in x and y direction within the transformed input image tie points and their manually marked conjugate points in reference TerraSAR-X image. Plot compares the alone and combined application of mutual information and SIFT operator for multimodal SAR image registration.

Keeping this in perspective, the combination of MI and SIFT seems to offer the following advantages:

i. The number of SIFT matches get definitely increase by using the initial estimate provided by MI.

ii. Once the matches have been found, these can be refined individually to fit an affine transformation which suffices for relatively large remote sensing scenes (not complete coverage's).

Table 6.12: Shift statistics (in meters) around the eight manually marked check points using different registration techniques.

	Absolute Average		Standard Deviation	
	X	Y	X	Y
MI	1.37	0.85	0.89	0.62
SIFT	1.60	2.90	1.30	1.60
SIFT + MI	1.20	1.90	1.02	0.98

The two registered images using SIFT and MI technique, is check-squared for visualization in Figure 6.19 and the overlay consistency of the on ground features like the airport runway is very high.

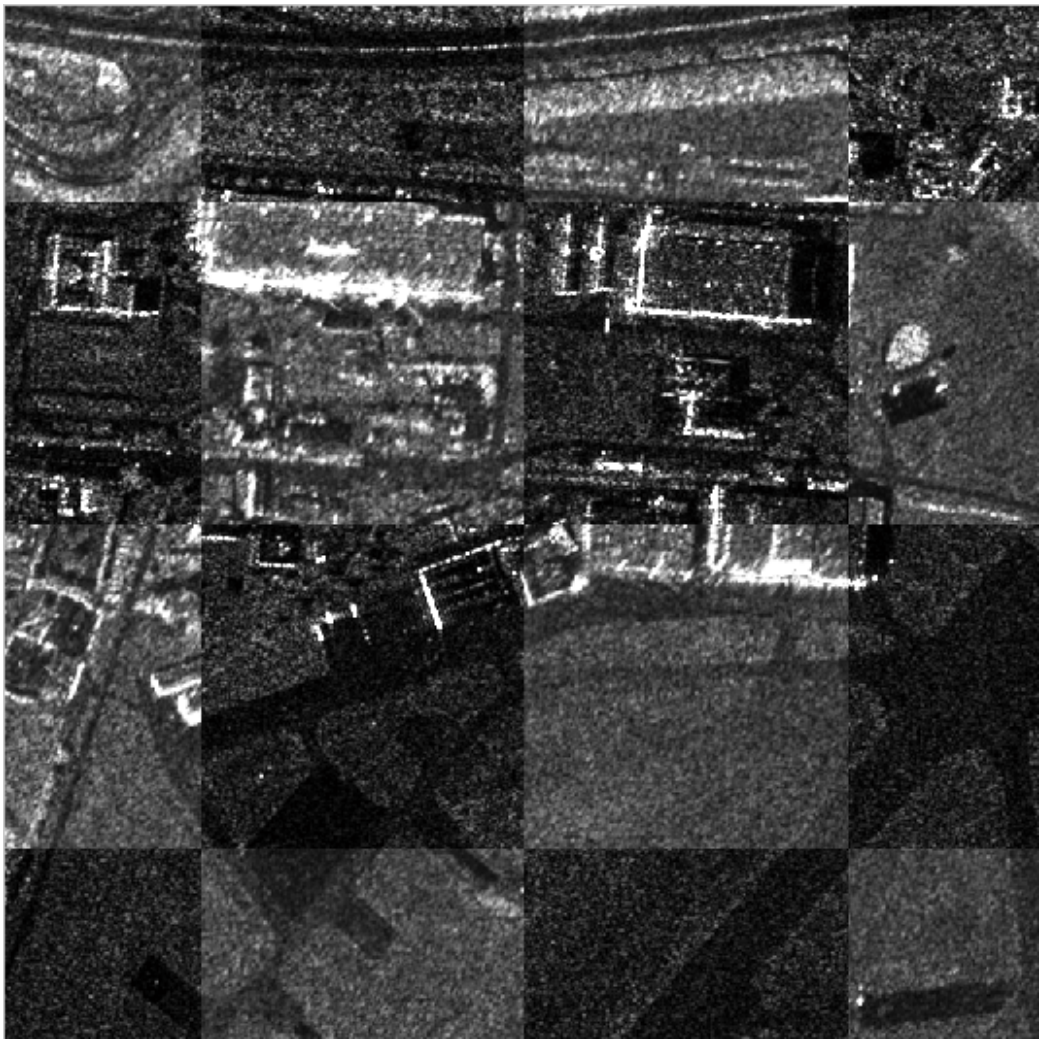


Figure 6.19: A small section of the registered TerraSAR-X (dark) and E-SAR imagery (bright). Images at 1 m pixel spacing shown without zoom.

6.10 CONCLUSION

We have analysed the performance of the SIFT operator for multimodal SAR image matching scenarios. For its specific adaptation to SAR imagery, we have considered changes in the feature detection, identification and matching stages of the original processing chain. Image smoothing of SAR images using ISEF filter and skipping features detected at the first octave increase the

operator matching efficiency and have reduced the processing time by almost 10 times for the tested 1000x1000 images. In the feature identification stage, we have compared the simple image differencing with two advanced gradient estimating techniques to show that the former technique performs at par with the latter much sophisticated approaches. It is also highlighted that getting rid of the rotation invariant characteristic of the SIFT can increase the number of matches considerably by keeping almost similar percentage accuracy. In general, it is found that the operator performance is optimum for same sensor images but still works consistently with sufficient accuracy for images with complex acquisition differences (like different sensors, incidence angles, orbit directions and speckle behaviour). To improve the SIFT performance with respect to the number of matches and their consistency, we utilize mutual information, to assist the operator matching scheme and if need be also for fine refinement to achieve desirable results.

CHAPTER 7

SUMMARY CONCLUSIONS AND PERSPECTIVE

The ever growing importance of satellite imagery in applications like treaty and border monitoring, disaster mitigation and many more, calls for continuous research and development in the field of remote sensing image processing. The reality that many remote sensing applications require fusion of multiple data sources (here satellite images) with different characteristics and advantages to achieve the desired objectives, asks for automated image matching and registration techniques. During the last decades, remote sensing sensors have undergone a rapid development in terms of both data quantity and characteristics. Due to these ever sprouting challenges, innovators are compelled to design new and/or improve existing remote sensing image registration techniques. Bringing to a close, we present as a summary, a chapter wise recall of all the major findings, conclusions and develop a perspective for the future, to fill in the gaps that came up during the course of this work.

In Chapter 1, we presented certain practical applications and necessity of completely automated image registration techniques. Examining the contemporary problems, we devised a problem statement and laid down our objectives to accomplish research in the field of multimodal image registration. The target imagery set were high resolution multimodal SAR and optical images from satellites like TerraSAR-X and IKONOS and we also emphasized the need of evolving a feature based matching technique for multimodal SAR images which are in vogue with the launch of high resolution satellites like TerraSAR-X, Tandem-X and Risat.

In Chapter 2, we discussed the basics of image registration and presented a thorough walk through of remote sensing image literature compiled from exclusive scientific journals and prestigious international conferences. We in detail presented advancements in intensity and feature based registration techniques and also covered certain aspects of image registration in the frequency domain. By establishing the “state of the art”, we eloquently highlighted the very

recent developments and challenges faced by the remote sensing community in multimodal image registration.

In Chapter 3, specifically catering to multimodal SAR and optical images, we presented a detailed discussion on intensity based registration techniques. For the elucidated task, highlighting the failure of Normalized Cross Correlation, we chose mutual information (MI) as the registration similarity metric. Discussing the origin of MI from information theory we examined all the nitty-gritty's of implementing MI based image registration. We introduced the application of higher order B-spline kernels in joint histogramming techniques showing them to be very robust but computationally intensive. Analysing the behaviour of simple to implement, fast in terms of execution time but ineffective for sub pixel image registration, we emphasized to use one step nearest neighbour technique (joint histogram estimation) especially for voluminous 1 m pixel spacing TerraSAR-X and IKONOS images. We proposed to utilize nearest neighbour with discrete optimization scenario (realized in Chapter 4) to achieve pixel level registration accuracy for high resolution images. This step was mainly targeted to reduce the execution time of intensity based registration which showed $O\left(n^2\right)$ behaviour, dependent on the image size, generally huge for remote sensing images. In this same chapter, we also advocated the effective use of image down sampling for MI based registration and showed limitations of mostly utilized simple translational shifts (x and y direction) as image registration parameters for relatively big Radarsat and Landsat images. Here, we also proposed to utilize MI locally within large remote sensing scenes to build a much effective affine transformation for image registration. An open concern still left here is to lay down a general guideline to choose between simple translation shifts and affine as the registration transformation function.

In Chapter 4, we specifically turned attention to high resolution SAR and optical imagery. We highlighted the impact of different acquisition principles of SAR and optical sensors on the response produced by real world 3D and 2D objects. Specifically, we analysed the behaviour especially of the SAR images in terms of information content and showed that pixels situated towards the higher end of a SAR image histogram do not contribute much to the image information content and can hamper the performance of MI as a registration similarity metric to achieve "on ground" registration. Analysing images acquired over sub urban areas we proposed a

histogram based segmentation scheme not only to improve MI performance in terms of registration accuracy but also in terms of processing time. As proposed in Chapter 3, we showed also effectiveness of discrete SPSA optimization (using simple nearest neighbour) to achieve pixel level registration consistency for high resolution SAR and optical image registration. However, still more tests and parameter study are required to use discrete optimization for high resolution images but we certainly believe that this is the way forward as in coming times the spatial resolution of the images will increase further (already WorldView-1 provides resolution of 50cm) quadrupling the data volume, produced by here analysed IKONOS images.

In Chapter 5, we focussed on a challenging problem to achieve image registration of high resolution SAR and optical imagery acquired over dense urban areas. We began our analysis by first highlighting the challenges of registering high resolution imagery of any modality. It was pointed out that achieving completely automatic image registration even within optical same sensor images might not be very straight forward. The influence of sensor geometry (decided by settings during scene acquisition) on appearance of real world objects was shown to be paramount both for same and different sensor images. Here, we extended the application of MI for registering SAR and optical images in urban areas. For this we presented two contrasting techniques with different pros and cons. The first technique is a more widely applicable histogram based solution (explored initially in Chapter 3) and the second is a more accurate technique based on extracted “on ground” regions, it is worthwhile mentioning that both the techniques required processing only in the reference SAR image to achieve desirable results. The performance of the two techniques here was evaluated using registration consistency measures and manual observations. For future applications, here we emphasized on developing a classification scheme for the reference image to accurately bin out “over ground” and “radar shadow” pixels participating in the registration process. This certainly does not look to be achieved comprehensively until now, though a very robust mutual information performance was reported. Based on the proposed histogram based solution, we presented a completely automated process to improve sensor orientation for orthorectification of high resolution optical data (scene acquired over the city of Munich, completely urban land cover). In this process, we utilized the geometric accuracy of the EEC TerraSAR-X products to establish ground control using local MI based matching for improving sensor orientation of optical data. Here, we also analysed, seldom

reported absolute registration accuracy achieved using MI. This was done by checking the orthorectification result with reference points estimated from a very accurate source. In general, it was found that MI has a capability to achieve pixel level absolute registration accuracy in absolutely plain areas and delivers accuracy of around 4-8 m for dense urban areas (for the Munich scene absolute accuracy was 4.5 m, but with undesirably low number of points participating in model refinement). This is definitely an improvement over the normally observed accuracy of IKONOS imagery (standard Geo product). We suggest this performance in urban areas can be further improved by incorporating an accurate classification scheme prior to the MI based registration process.

In Chapter 6, we engaged efforts in evolving a computer vision technique Scale Invariant Feature Transform (SIFT) for remote sensing SAR images. Understanding the possible incompatibilities of the originally proposed SIFT processing chain in all feature detection, identification and matching stages, we analysed changes to enhance the operator performance for microwave remote sensing applications. Starting with feature detection, we proposed image smoothing using ISEF filter and also skipping features detected at the first octave to improve the processing time by an approximate factor of ten for the tested 1000x1000 pixel images. The introduced steps help the image differencing based SIFT feature detection perform very nicely for SAR images with inherent multiplicative speckle influence. Turning attention to feature identification, we compared the descriptors formed using simple image differencing with much sophisticated techniques like recursive image differencing and image ratio (both specifically designed for SAR edge detection). It was noted that the descriptors formed using simple image differencing performed at par to the two selected alternatives. Incapability, specifically of descriptors formed using ratio based operator ROEWA for images having a rotation difference was observed mainly due to narrow range of edge orientations arising due to the basic nature of image ratio. Developing a method to assign signs to the x and y gradients (to broaden the edge orientation range and thus generating much robust feature descriptors) estimated using ratio operators can be considered as a future challenge in this area, as gradients estimated using ROEWA do show slightly higher match accuracy (percentage wise). However, it was found that number of matches can be increased drastically by giving away the rotation invariance characteristic of the SIFT feature descriptors, which is generally acceptable for already rectified

satellite data. From the observation that the modified SIFT detection scheme show high number of feature repeatability but comparatively fewer number of successful matches, we investigated the effects of assisting the SIFT standard matching with MI. This proposed hybrid chain showed remarkable improvement over the SIFT processing chain for a challenging multitemporal, multisensor SAR dataset acquired over a semi urban land cover. Ideally, such a hybrid chain combining benefits of a robust feature detector and a similarity metric should be realised for SAR and optical image registration as well, but due to disparate sensor characteristics, this looks a far distant reality.

To conclude, according to the laid down objectives in Chapter 1, this dissertation has contributed the following in the field of multimodal remote sensing image registration:

i. With two proposed methods, we have presented implementations of MI suited to high resolution SAR and optical imagery acquired especially over urban areas.
ii. On basis of the improved MI performance, we have presented a novel approach to increase the absolute accuracy of high resolution optical data by improving the sensor orientation.
iii. The performance of the SIFT operator, originally proposed for digital camera images, has been refined and extended for multimodal SAR image matching and registration applications.

REFERENCES

Bamler, R., and Eineder, M., 2009. The pyramids of Gizeh seen by TerraSAR-X-a prime example for unexpected scattering in SAR, *IEEE Geoscience and Remote Sensing Letters*, 5(3), pp. 468-470

Battiato, S., Gallo, G., Puglisi, G., and Scellato, S., 2007. SIFT features tracking for video stabilization, *in Proc. 14th International Conference on Image Analysis and Processing*, pp. 825-830, Modena, Italy, Sep 2007

Bay, H., Ess, A., Tuytelaars, T. and Gool, L.V., 2008. Speeded-Up Robust Features (SURF), *Computer Vision and Image Understanding*, 110, pp. 346-359

Beis, J., and Lowe, D.G., 1997. Shape indexing using approximate nearest neighbour search in high dimensional spaces, *in Proc. IEEE Conference on Computer Vision Pattern Recognition*, pp. 1000-1006, June 1997

Bentoutou, Y., Taleb, N., Kpalma, K. and Ronsin, J., 2005. An automatic image registration for applications in remote sensing, *IEEE Transactions on Geoscience and Remote Sensing*, 43, pp. 2127-2137

Bhattacharya, M., and Das, A., 2007. Multi resolution medical image registration using maximization of mutual information and optimization by genetic algorithm, *in Proc. IEEE Nuclear Science Symposium, Medical Imaging Conference*, 4, pp. 2961-2964, Honolulu, Hawaii, USA, Oct.28-Nov.3 2007

Bookstein, F.L., 1989. Principal warps: thin-plate splines and the decomposition of deformations, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, pp. 567-585

Borghys, D., Perneel, C., and Acheroy, M., 2001. A hierarchical approach for registration of high-resolution polarimetric SAR images, *in Proc. SPIE Conference on Image and Signal Processing for Remote Sensing*, Toulouse, France, Sep 2001

Bovik, A.C., 1988. On detecting edges in speckle imagery, *IEEE Transactions on Acoustics*, *Speech and Signal Processing*, 36(10), pp. 1618-1627

Bovolo, F., and Bruzzone, L., 2005. A detail-preserving scale-driven approach to change detection in multitemporal SAR images, *IEEE Transactions on Geoscience and Remote Sensing*, 43(12), pp. 2963-2972

Bracewell, R., 1965. The Fourier transform and its applications, McGraw-Hill, New York

Brooks, R., and Bovik, A., 1990. Robust techniques for edge detection in multiplicative Weibull image noise, *Pattern Recognition*, 23(10), pp. 1047-1057

Brown, L.G., 1992. A survey of image registration techniques, *ACM Computing Surveys*, 24 (4), pp. 325-376

Canny, J.F., 1983. Finding edges and lines in images, *Massachusetts Institute of Technology Technical Report*, 720

Canny, J.F., 1986. A computational approach to edge detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6), pp. 679-698

Castleman, K.R., 1996. Digital image processing, Upper Saddle River, EUA: Prentice-Hall

Chaniotis, A.K., Poulikakos, D., 2004. High order interpolation and differentiation using B-splines, *Journal of Computational Physics*, 197, 1, pp.253-274

Chen, H., and Varshney, P.K., 2000. A pyramid approach for multimodality image registration based on mutual information, *in Proc. 3rd International Conference on Information Fusion,* 1, pp. 9-15, Paris, France, July 2000

Chen, H., and Varshney, P.K., 2003. Mutual information-based CT-MR brain image registration using generalized partial volume joint histogram estimation, *IEEE Transactions on Medical Imaging*, 22, pp. 1111-1119

Chen, H., and Varshney, P.K., 2004a. MI based registration of multisensor and multi-temporal images, *in: P. K. Varshney and M. K. Arora* (eds.), *Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data*, Springer Verlag, 2004.

Chen, H., and Varshney, P.K., 2004b. Size-dependent image resampling for mutual information based remote sensing image registration, *in Proc. IEEE International Geoscience and Remote Sensing Symposium*, Anchorage, Alaska, Sep 2004

Chen, H., Varshney, P.K., and Arora, M.K., 2003a. Mutual information based image registration for remote sensing data, *International Journal of Remote Sensing,* 24(18), pp. 3701-3706

Chen, H., Varshney, P.K., and Arora, M.K., 2003b. Performance of mutual information similarity measure for registration of multitemporal remote sensing images, *IEEE Transactions on Geoscience and Remote Sensing*, 41(11), pp. 2445-2454

Cheng, H., Zheng, S., Yu, Q., Tian, J., Liu, J., 2004. Matching of SAR images and optical images based on edge feature extracted via SVM, *in Proc. 7th International Conference on Signal Processing*, pp. 930-933, Beijing, China, Aug 2004

Cole-Rhodes, A.A., Johnson, K.L., Moigne, J.,Le, and Zavorin, I., 2003a. Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient, *IEEE Transactions on Image Processing*, 12(12), pp. 1495-1511

Cole-Rhodes, A.A., Johnson, K.L., and Moigne, J.,Le, 2003b. Image registration using a 2nd order stochastic optimization of mutual information, *in Proc. Geoscience and Remote Sensing Symposium,* pp. 4038-4040, Toulouse, France, July 2003

Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., Marchal, G., 1995. Automated multi-modality image registration based on information theory, *Information Processing in Medical Imaging*, pp. 263-274

Cover, T.M., and Thomas, J.A., 1991. Elements of information theory, John Wiley and Sons, Inc

Cover, T.M., and Thomas, J.A., 2006. Elements of information theory, second edition, John Wiley and Sons, Inc

Dai, X., and Khorram, S., 1998. The effects of image misregistration on the accuracy of remotely sensed change detection, *IEEE Transactions on Geoscience and Remote Sensing*, 36(5), pp. 1556-1577

Dare, P., and Dowman, I., 2000. An improved model for automatic feature-based registration of SAR and SPOT images, *ISPRS Journal of Photogrammetry and Remote Sensing*, 56, 2000

Dasgupta, D., and McGregor, D.R., 1992. Digital image registration using structured genetic algorithms, *in Proc. SPIE the International Society for Optical Engineering*, 1766, pp. 226-234.

De Castro, E., and Morandi, C., 1987. Registration of translated and rotated images using finite Fourier transforms, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5), pp. 700-703

Deriche, R., 1987. Using Canny's criteria to derive a recursively implemented optimal edge detector, *International Journal of Computer Vision*, 1(2), pp. 167-187

Devereux, B.J., Fuller, R.M., Carter, L., and Parsell, R.J., 1990. Geometric correction of airborne scanner imagery by matching Delaunay triangles, *International Journal of Remote Sensing*, 11(12), pp. 2237-2251

Djamdji, J.P., Bijaoui, A., and Mamère, R., 1993. Geometrical registration of images, the multiresolution approach, *Photogrammetric Engineering and Remote Sensing*, 59, pp. 645-653

Dupont, S., Nonin, P., Renouard, L., Pichon, G., and Berthod, M., 1997, ERS SAR interferometry: an operational evaluation of the DTM production, *Remote Sensing - A Scientific Vision for Sustainable Development*, 3, pp. 1347-1349

Eineder, M., Breit, H., Fritz, T., Schättler, B., and Roth, A., 2005. TerraSAR-X SAR products and processing algorithms, *in Proc. IEEE International Geoscience and Remote Sensing Symposium*, 7, pp. 4870-4873, Seoul, Korea, July 2005

El-ghazawi, A.T., Chalermwat, P., and Moigne, J.,Le, 1997. Wavelet based image registration on parallel computers, *in Proc. ACM/IEEE Supercomputing Conference*, San Jose, California, USA, Nov 1997

Fiala, M., and Shu, C., 2006. 3D model creation using self-identifying marks and SIFT features, *IEEE Workshop on Haptic Audio Visual Environments and their Applications* (HAVE'06), Ottawa, Ontario, Nov 2006

Fischler, M.A., and Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Communication of the ACM*, 24(6), pp. 381-395

Fjørtoft, R., Lopes, A., Bruniquel, J., and Marthon, P., 1999. Optimal edge detection and edge localization in complex SAR images with correlated speckle, *IEEE Transactions on Geoscience and Remote Sensing*, 37(5), pp. 2272-2281

Fjørtoft, R., Marthon, P., Lopes, A., and Cubero-Castan, E., 1995. Edge detection in radar images using recursive filters, *in Proc. ACCV'95*, III, pp. 87-91, Singapore, Dec 1995

Flusser, J., 1992. An adaptive method for image registration, *Pattern Recognition*, 25, pp. 45-54

Fonseca, L.M.G., and Manjunath, B.S., 1996. Registration techniques for multisensor remotely sensed imagery, *Photogrammetric Engineering and Remote Sensing*, 62(9), pp. 1049-1056

Foroosh, H., Zerubia J.B., Berthod, M., 2002. Extension of phase correlation to subpixel registration, *IEEE Transaction on Image Processing*, 11(3), pp. 188-200

Förstner, W., and Gülch, E., 1987. A fast operator for detection and precise location of distinct points, corners and centres of circular features, *in Proc. ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, pp. 281-305

Fritz, T., and Eineder, M., 2008. TerraSAR-X ground segment, *SAR basic product specification document (TX-GS-DD-3302)*, Issue 1.5, February 2008

Frost, V., Stiles, J., Shanmugan, K., and Holtzman, J., 1982. A model for radar images and its applications to adaptive digital filtering of multiplicative noise, *IEEE Transactions on Pattern Analysis and Computer Intelligence*, 4(2), pp. 157-165

Goshtasby, A., 1985. Template matching in rotated images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7, pp. 338-344

Goshtasby, A., 1986. Piecewise linear mapping functions for image registration, *Pattern Recognition*, 19, pp. 459-466

Goshtasby, A., 1987. Piecewise cubic mapping functions for image registration, *Pattern Recognition,* 20, pp. 525-533

Goshtasby, A., 1988a. Image registration by local approximation methods, *Image and Vision Computing*, 6, pp. 255-261

Goshtasby, A., 1988b. Registration of images with geometric distortions, *IEEE Transactions on Geoscience and Remote Sensing*, 26, pp. 60-64

Goshtasby, A., Stockman, G., Rohr, K., 2004. Tutorial: 2-D and 3-D image registration, *IEEE Conference on Computer Vision and Pattern Recognition*, Washington D.C USA. Available at: http://www.imgfsr.com/IntroductionCVPR.pdf

Goshtasby, A., Stockman, G.C., 1985. Point pattern matching using convex hull edges, *IEEE Transactions on Systems, Man and Cybernetics*, 15, pp. 631-637

Goshtasby, A., Stockman, G.C., Page, C.V., 1986. A region-based approach to digital image registration with subpixel accuracy, *IEEE Transactions on Geoscience and Remote Sensing*, 24, pp. 390-399

Grodecki, J., and Dial, G., 2001. IKONOS geometric accuracy, Available at http://www.satimagingcorp.com/satellite-sensors/IKONOSGeometricAccuracy-ISPRS202001.pdf

Grün, A., 1985. Adaptive least squares correlation - a powerful image matching technique, *South African Journal of Photogrammetry, Remote Sensing and Cartography*, 4(3), pp. 175-187

Hardy, R.L., 1990. Theory and applications of the multiquadric-biharmonic method, Computers *and Mathematical Applications* 19 (8/9), pp. 163-208

Harris, C.G., and Stephens, M., 1988. A combined corner and edge detector, *in Proc. 4th Alvey Vision Conference.*, pp. 147-151, Manchester, UK, Sep 1988

Hegart-Macle, L.S., Zribi, M., Alem, F., Weisse, A., and Loumagne, C., 2002. Soil moisture estimation from ERS/SAR data towards an operational methodology, *IEEE Transactions on Geoscience and Remote Sensing*, 40, pp. 2647-2658

Hill, S.D., 2005. Discrete stochastic approximation with application to resource allocation, *in the Johns Hopkins APL Technical Digest*, Jan - Mar 2005, 26 (1), pp. 15-21

Hoja, D., Reinartz, P., Schroeder, M., 2006. Comparison of DEM generation methods using high resolution optical stereo imagery and interferometric SAR data, in Proc. *ISPRS Commission I Symposium 2006 From Sensors to Imagery*, Paris, France, Vol. XXXVI, Part 1/B

Holden, M., Hill, D.L.G., Denton, E.R.E., Jarosz, J.M., Cox, T.C.S., Rohlfing, T., Goodey, J., Hawkes, D. J., 2000. Voxel similarity measures for 3-D serial MR brain image registration, *IEEE Transactions on Medical Imaging*, 19(2), pp. 94-102

Hough, P.V.C., 1992. Methods and means for recognizing complex patterns. U.S. Patent 3069654

Hu, M.K., 1962. Visual pattern recognition by moment invariants, *IEEE Transactions on Information Theory*, 8, pp. 179-187

Inglada, J., 2002. Similarity measures for multi sensor remote sensing images, *in Proc. IEEE Geoscience and Remote Sensing Symposium*, Toronto, Canada, July 2002

Inglada, J., Adragna, F., 2001. Automatic multi-sensor image registration by edge matching using genetic algorithms, *in Proc. IEEE Geoscience and Remote Sensing Symposium*, 5, pp. 2313-2315, Sydney, Australia, July 2001

Inglada, J., and Giros, A., 2004. On the possibility of automatic multisensor image registration, *IEEE Transactions on Geoscience and Remote Sensing*, 42(10), pp. 2104-2120

Inglada, J., Muron, V., Pichard, D., and Feuvrier, T., 2007. Analysis of artefacts in subpixel remote sensing image registration, *IEEE Transactions on Geoscience and Remote Sensing*, 45(1), pp. 254-264

Jacobson, K., (2004). Direct / integrated sensor orientation - pros and cons, *in Proc. ISPRS Congress, Istanbul 2004*. Invited paper. Available at: http://www.isprs.org/congresses/istanbul2004/comm3/papers/384.pdf

Kanal, L.N., Lambird, B.A., Lavine, D., and Stockman, G.C., 1981. Digital registration of images from similar and dissimilar sensors, *in Proc. International Conference on Cyberbetics and Society*, pp. 347-351

Keys, R.G., 1981. Cubic convolution interpolation for digital image processing, *IEEE Transactions on Acoustics, Speech and Signal Processing*, 29(6), pp. 1153-1160

Krieger, G., Fiedler, H., Zink, M., Hajnsek, I., Younis, M., Huber, S., Bachmann, M., Hueso Gonzalez, J., Werner, M. and Moreira, A., 2007. The TanDEM-X mission: A satellite formation for high resolution SAR interferometry, *in Proc. International Astronautical Congress,* Hyderabad, India*,* Sep 2007

Lavine, D., Lambird, B., and Kanal, L., 1983. Recognition of spatial point patterns, *Pattern Recognition*, 16, pp. 289-295

Lehmann, T.M., Onner, C.G., and Spitzer, K., 1999. Survey: interpolation methods in medical image processing, *IEEE Transactions on Medical Imaging*, 18, pp. 1049-1075

Lehureau, G., Tupin, F., Tison, C., Oller, G., and Petit, D., 2008. Registration of metric resolution SAR and optical images in urban areas, *in Proc. 7th European Conference on Synthetic Aperture Radar*, Friedrichshafen, Germany, June 2008

Lewis, J.P., 1995. Fast normalized cross-correlation, *Vision Interface*, pp. 120-123. Canadian Image Processing and Pattern Recognition Society

Li, H., Manjunath, B.S., and Mitra, S.K., 1995. A contour-based approach to multisensor image registration, *IEEE Transactions on Image Processing*, 4(3), pp. 320-334

Li, H.H., and Zhou, Y.T., 1996. A wavelet based point feature extractor for multi sensor image registration, *SPIE Journal of Optical Engineering*, 35(2), pp. 391-400

Li, W., Li, X., Wu, Y., Hu, Z., 2006. A novel framework for urban change detection using VHR satellite images, *Pattern Recognition*, 2, pp. 312-315

Lindeberg, T., and Garding, J., 1997. Shape-adapted smoothing in estimation of 3-D depth cues from affine distortions of local 2-D brightness structure, *Image and Vision Computing*, 15, pp. 415-434

Lindeberg, T., 1998. Feature detection with automatic scale selection, *International Journal of Computer Vision*, 30, pp. 77-116

Lowe, D.G., 1999. Object recognition from local scale-invariant features, *in Proc. International Conference on Computer Vision*, pp. 1150-1157, Corfu, Greece, Sep 1999

Lowe, D.G., 2004. Distinctive image features from scale-invariant features, *International Journal of Computer Vision*, 60(2), pp. 91-110

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P., 1997. Multimodality image registration by maximization of mutual information, *IEEE Transactions on Medical Imaging*, 16, pp. 187-98

Maintz, J.B.A., and Viergever, M.A., 1998. A survey of medical image registration, *Medical image analysis, 2*, pp. 1-36

Marr, D., and Hilderth, E., 1980. Theory of edge detection, *in Proc. Royal Society of London*, B 207, pp. 187-217

Mikolajczyk, K., and Schmid, C., 2001. Indexing based on scale invariant interest points, *in Proc. of the 8th International Conference on Computer Vision*, Vancouver, Canada, July 2001

Mikolajczyk, K., and Schmid, C., 2004. Scale and affine invariant interest point detectors, *International Journal of Computer Vision*, 60, pp. 63-86

Mikolajczyk, K., and Schmid, C., 2005. A performance evaluation of local descriptors, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27, pp. 1615-1630

Moghaddam, M., Dungan, J.L., and Acker, S., 2002. Forest variable estimation from fusion of SAR and multispectral optical data, *IEEE Transactions on Geoscience and Remote Sensing*, 40, pp. 2176-2187

Moigne, J.,Le, Netanyahu, N., Masek, J., Mount, D., Goward, S., 2001.Robust matching of wavelet features for sub-pixel registration of Landsat data, *in Proc. IEEE Geoscience and Remote Sensing Symposium,* Sydney, Australia, July 2001

Moigne, J.,Le, 1995. Towards a parallel registration of multiple resolution remote sensing data, *in Proc. IEEE Geoscience and Remote Sensing Symposium*, Florence, Italy, July 1995

Nonaka, T., Ishizuka, Y., Yamane, N., Shibayama, T., 2008. Evaluation of the geometric accuracy of TerraSAR-X, *in Proc. ISPRS Congress*, Beijing, China, July 2008

Oller, G., Marthon, P., Denise, L., 2002. SAR image matching using the edge strength map, *in Proc. IEEE Geoscience and Remote Sensing Symposium*, Toronto, Canada, June 2002

Oller, G., Petit, D., and Inglada, J., 2006. On the use of SAR and optical images combination for scene interpretation, *in Proc. ISPRS Symposium, Commission I, WGI/3, From Sensors to Imagery*, Paris, France, 2006

Paillau, P., 1997. Detecting step edges in noisy SAR images, a new linear operator, *IEEE Transactions on Geoscience and Remote Sensing,* 35, pp. 191-196

Palubinskas, G., Meyer, F.J., Runge, H., Reinartz, P., Scheiber, R., and Bamler, R., 2005. Estimation of along-track velocity of road vehicles in SAR data, *in Proc. of SPIE*, 5982, pp. 1-9

Panigrahi, N., and Tripathy, S., 2002. Image registration using polynomial affine transformation, *Defence Science Journal*, 52(3), pp. 253-259

Parker, J.A., Kenyon, R.V., and Troxel, D.E., 1983. Comparison of interpolating methods for image resampling, *IEEE Transactions on Medical Imaging*, 2, pp. 31-39

Pluim, J.P.W., Maintz, I.B.A., Viergever, M., 2000. Interpolation artefacts in mutual information based image registration, *Computer Vision and Image Understanding*, 77, pp. 211-232

Pluim, J.P.W., Maintz, I.B.A., Viergever, M., 2003. Mutual information based registration of medical images: a survey, *IEEE Transactions on Medical Imaging*, 22(8), pp. 986-1004

Pohl, C., and van Genderen, J.L., 1998. Multisensor image fusion in remote sensing: concepts, methods and applications, *International Journal of Remote Sensing*, 19, pp. 823-854

Ranade, S., and Rosenfeld, A., 1980. Point pattern matching by relaxation, *Pattern Recognition*, 12, pp. 269-275

Reddy, B.S., and Chatterji, B.N., 1996. An FFT-based technique for translation, rotation, and scale-invariant image registration, *IEEE Transactions on Image Processing*, 5(8), pp. 1266-1271

Reinartz, P., Lachaise, M., Schmeer, E., Krauss, T., Runge, H., 2006a. Traffic monitoring with serial images from airborne cameras, *ISPRS Journal of Photogrammetry and Remote Sensing*, 61, 3-4, pp. 149-158

Reinartz, P., Müller, R., Lehner, M., Schroeder, M., 2006b. Accuracy analysis for DSM and orthoimages derived from SPOT HRS stereo data using direct georeferencing, *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(3), pp. 160-169

Reinartz, P., Müller, R., Suri, S., Schwind, P., Bamler, R., 2009. Using geometric accuracy of Terrasar-X data for improvement of direct sensor orientation and ortho-rectification of VHR optical sensor data, *in Proc. IEEE International Geoscience & Remote Sensing Symposium*, Cape Town, South Africa, July 2009

Roche, A., Pennec, X., Rudolph, M., Auer, D.P., Mal, G., Ourselin, S., Auer, L.M., Ayache, N., 2000. Generalized correlation ratio for rigid registration of 3D ultrasound with MR images, *in Proc. of MICCAI'00, LNCS 1935,* pp. 567-577, 2000

Rohr, K., Stiehl, H.S., Buzug, T.M., Weese, J., Kuhn, M.H., 2001. Landmark based elastic registration using approximating thin-plate splines, *IEEE Transactions on Medical Imaging*, 20, pp.526-534

Roth, A., 2003. TerraSAR-X: A new perspective for scientific use of high resolution spaceborne SAR data, *in Proc. 2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, Berlin, Germany, May 2003

Schwind, P., 2008. Critical evaluation of the SIFT operator for Remote Sensing Imagery, *Diplomarbeit*, Remote Sensing Technology Institute (IMF), German Aerospace Centre (DLR)

Schwind, P., Suri, S., Reinartz, P., Siebert, A., in print. Applicability of the SIFT operator for geometrical SAR image registration. Accepted in *International Journal of Remote Sensing*

Schreiber, R., Reigber, A., Ulbricht, A., Papathanassiou, K., Horn, R., Buckreuss, S. and Moreira, A., 1999. Overview of interferometric data acquisition and processing modes of the experimental airborne SAR system of DLR, *in Proc. IEEE International Geoscience and Remote Sensing Symposium*, Hamburg, Germany, July 1999

Shannon, C.E., 1948. A mathematical theory of communication, *The Bell System Technical Journal*, 27, pp. 379-423, 623-656. Available at http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf

Shaoyan, S., and Chonghui, G., 2007. Medical image registration by maximizing a hybrid normalized mutual information, *in Proc. 1st International Conference on Bioinformatics and Biomedical Engineering*, pp. 964-967, Wuhan University, China, July 2007

Shen, J., and Castan, S., 1986. An optimal linear operator for edge detection, *in Proc. IEEE International Conference on Computer Vision and Pattern Recognition,* Miami, Florida, 1986

Shen, J., and Castan, S., 1992. An optimal linear operator for step edge detection, *CVGIP: Graphical Models and Image Processing*, 54(12), pp. 112-133

Shu, L., Tan, T., Tang, M., and Pan, C., 2005. A novel registration method for SAR and SPOT images, *in Proc. International Conference on Image Processing* , pp. 213-216, Genoa, Italy, Sep 2005

Simoncelli, E.P., and Freeman, W.T., 1995. The Steerable Pyramid: a flexible architecture for multi-scale derivative computation, *in Proc. IEEE Second International Conference on Image Processing*, Washington DC, Oct 1995

Spall, J.C., 1992. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation, *IEEE Transactions on Automatic Control*, 37, pp. 332-341

Spall, J.C., 1998. An overview of the simultaneous perturbation method for efficient optimization, *in the Johns Hopkins APL Technical Digest*, 19(4). Available at http://techdigest.jhuapl.edu/td1904/spall.pdf

Stilla, U., 2007. High resolution radar imaging of urban areas, *in: Fritsch D (ed) Photogrammetric Week, 2007*, Wichmann, Heidelberg, pp. 149-158 (invited paper). Available at: http://www.ifp.uni-stuttgart.de/publications/phowo07/180Stilla.pdf

Stockman, G., Kopstein, S., and Benett, S., 1982. Matching images to models for registration and object detection via clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 4, pp. 229-241

Stone, H.S., 1998. Fast correlation techniques for wavelet representations, *in: Proc. SPIE Conference on AeroSense, Aerospace/Defense Sensing and Controls*, Orlando, Florida USA, April 1998

Stone, H.S., Tao, B., and McGuire, M., 2003. Analysis of image registration noise due to rotationally dependent aliasing, *Journal of Visual Communication and Image Representation*, 14(2), pp. 114-135

Stramondo, S., Bignami, C., Pierdicca, N., Chini, M., 2007. SAR and optical remote sensing for urban damage detection and mapping: case studies, *in Proc. Urban Remote Sensing Joint Event*, pp. 1-6, Paris, France, April 2007

Studholme, C., Hill, D. L. G., Hawkes, D. J., 1999. An overlap invariant entropy measure of 3D medical image alignment, *Pattern Recognition*, 32(1), pp. 71-86

Sun, S., Guo, C., 2007. Image registration by minimizing Tsallis divergence measure, *in Proc. 4th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)*, 4, pp.712-715, 2007, Hainan, China, Aug 2007

Suri, S., and Reinartz, P., 2008. Application of generalized partial estimation for mutual information based registration of high resolution SAR and optical imagery, *in Proc: 11th International Conference on Information Fusion*, Cologne Germany, June 30- July 3, 2008

Suri, S., and Reinartz, P., 2009. On the possibility of intensity based registration for metric resolution imagery, *in Proc. 12th International AGILE Conference on Geographic Information Science*, June 2009, Hannover, Germany

Suri, S., and Reinartz, P., 2010. Mutual information based registration of TerraSAR-X and IKONOS imagery in urban area, *IEEE Transactions on Geoscience and Remote Sensing*, 48(2), pp. 939-949

Suri, S., Schwind, P., Reinartz, P., and Uhl, J., 2009a. Combining mutual information and scale invariant feature transform for fast and robust multisensor SAR image registration, *in Proc. 75th ASPRS Annual Conference*, Baltimore MD, USA, March 2009

Suri, S., Türmer, S., Reinartz, P., and Stilla, U., 2009b. Registration of high resolution SAR and optical imagery in urban areas, *in Proc. ISPRS High-Resolution Earth Imaging for Geospatial Information*, June 2009, Hannover, Germany

Thévenaz, P., and Unser, M., 2000. Optimization of mutual information for multiresolution image registration, *IEEE Transactions on Geoscience and Remote Sensing*, 41(11), pp. 2445-2454

Thévenaz, P., Blu, T., and Unser, M., 2000. Interpolation revisited, *IEEE Transactions on Medical Imaging*, 19, pp. 739-758

Touzi, R., Lopes, A., and Bousquet, P., 1988. A statistical and geometrical edge detector for SAR images, *IEEE Transactions on Geoscience and Remote Sensing*, 26, pp.764-773

Tsao, J., 2003. Interpolation artefacts in multimodality image registration based on maximization of mutual information, *IEEE Transactions on Medical Imaging*, 22(7), pp. 854-864

Türmer, S., 2009. Automatic registration of high resolution SAR and optical imagery in urban areas, *Diplomarbeit*, Remote Sensing Technology Institute (IMF), German Aerospace Centre (DLR)

Tupin, F., and Nicolas, J.M., 2002. Matching criteria for radargrammetry, *in Proc. IEEE Geoscience and Remote Sensing Symposium*, Toronto, Canada, July 2002

Turton, B., Arslan, T., and Horrocks, D., 1994. A hardware architecture for a parallel genetic algorithm for image registration, *in Proc. IEEE Colloquium on Genetic Algorithms in Image Processing and Vision,* pp. 11/1-11/6, London, UK, 1994

Viola, P.A., 1995, Alignment by maximization of mutual information, *PhD Thesis*, Artificial Intelligence Laboratory, Massachusetts Institute of Technology Source: http://research.microsoft.com/~viola/Pubs/MIT/PHD-thesis.pdf

Wachowiak, M.P., Smolikova, R., and Peters, T.M., 2003. Multiresolution biomedical image registration using generalized information measures, *Lecture Notes in Computer Science,* 2879, (MICCAI 2003), pp. 846-853

Wegner, J.D., and Soergel, U., 2008. Registration of SAR and optical images containing Bridges over land, *in Proc. EARSel Workshop Remote Sensing - New Challenges of High Resolution*, Bochum, Germany, March 2008

Wegner, J.D., Inglada, J., and Tison, C., 2008. Automatic fusion of SAR and optical imagery based on line features, *in Proc. 7th European Conference on Synthetic Aperture Radar*, Friedrichshafen, Germany, June 2008

Wong, A., and Clausi, D.A., 2007. ARRSI: Automatic registration of remote sensing images, *IEEE Transactions on Geoscience and Remote Sensing*, 45 (5) Part II, pp. 1483-1493

Woods, R.P., Cheery, S.R., and Mazziotta, J.C., 1992. Rapid automated algorithm for aligning and reclining PET images, *Journal of Computer Assisted Tomography*, 16(4), pp. 620-633

Woods, R.P., Grafton, S.T., Holmes, C.J., Cheery, S.R., Mazziotta, J.C., 1998. Automated image registration: general methods and intrasubject, intramodality validation, *Journal of Computer Assisted Tomography*, 22(1), pp.139-152

Xie, H., Pierce, L.E., and Ulaby, F.T., 2003. Mutual information based registration of SAR images. In *Proc. IEEE Geoscience and Remote Sensing Symposium*, 6, pp.4028- 4031, Toulouse, France, 28 June - 02 July 2003

Xie, J., Chen, Z., and Xu, G., 2008. A new method for medical image registration based on wavelet multi-scale transform, *in Proc. ISECS International Colloquium on Computing, Communication, Control, and Management,* Guangzhou City, China, Aug 2008

Xin, Y., Li, J., and Cheng, Q., 2007. Automatic generation of remote sensing image mosaics for mapping large natural hazards areas, *Lecture Notes in Geoinformation and Cartography*, *Geomatics Solutions for Disaster Management*, Springer Berlin Heidelberg, pp. 61-73

Yang, W., Han, C., Sun, H., Cao, C., 2005. Registration of high resolution SAR and optical images based on multiple features, *in Proc. IEEE Geoscience and Remote Sensing Symposium*, Seoul, Korea, July 2005

Zibaeifard, M., and Rahmati, M., 2006. An adaptive simulated annealing scheme for multi-modality medical image registration by maximization of mutual information, *in Proc. 8th International Conference on Signal Processing*, Beijing, China, Nov 2006

Zitová, B., and Flusser, J., 2003. Image registration methods: a survey, *Image and Vision Computing*, 21(11), pp. 977-1000

APPENDIX

SCALE INVARIANT FEATURE TRANSFORM (SIFT)

1.0 INTRODUCTION

The SIFT operator is a computer vision algorithm first introduced by David Lowe from university of British Columbia, Canada (Lowe, 1999). The method developed for optical camera images provides a robust technique to extract and match distinctive features within images with various acquisition differences. Moreover, SIFT features are invariant to image scale and rotation, which makes them well suited to match different views of an object or scene. Later, Lowe published a revised version of the SIFT operator with several improvements in terms of feature accuracy, stability and invariance (Lowe, 2004). Since its introduction it has proven its effectiveness for numerous applications like object recognition, robot localization and mapping and panorama stitching in the field of computer vision. In this appendix, the originally proposed SIFT operator for optical camera images is detailed here for a clear understanding of the results presented in Chapter 6.

2.0 SIFT PROCESSING CHAIN

The SIFT processing chain has four major processing stages:

i. **Scale space extrema detection**: Using the Difference of Gaussians function, a scale space is built, out of which extrema are detected.
ii. **Feature localization**: The found extrema are filtered by different criteria to select only features with high stability.
iii. **Orientation assignment**: By analysing the local image gradient directions, at least one orientation is assigned to every feature.
iv. **Feature descriptor**: Making use of the location, scale and rotation of the features, descriptors are created.

2.1 Scale Space Extrema

Scale space

The first step of the SIFT operator is to build a scale space for each of the involved images. The scale space $L(x;y;\sigma)$ is defined as an input image $I(x,y)$ that has been convolved several times with a two-dimensional Gaussian function $G(x;y;\sigma)$ using an increasing standard deviation σ

$$L(x,y,\sigma) = I(x,y) * G(x,y,\sigma) \tag{1}$$

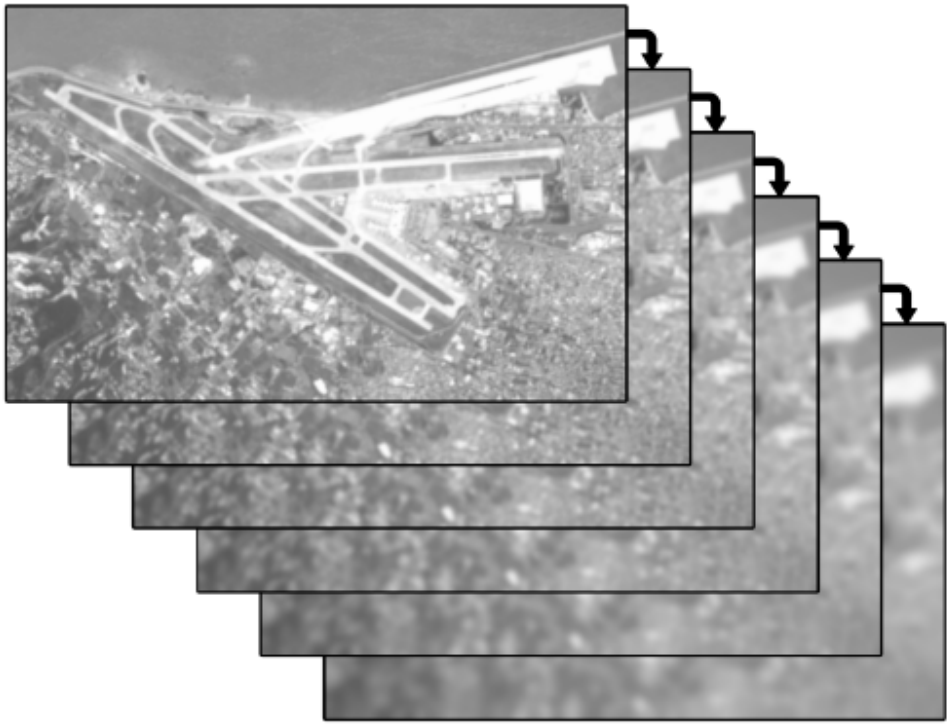


Figure 1: An octave of the scale space, depicting an optical image filtered repeatedly with Gaussian filter of increasing standard deviation. (Scene area: Beirut Airport, Lebanon) (Schwind, 2008).

where the Gaussian function is defined as

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2}\exp^{\frac{-(x^2+y^2)}{2\sigma^2}} \tag{2}$$

The purpose of this scale space is to allow the detection of features at different scales. This is possible as more and more of the smaller structures get blurred out every time the image is Gauss

filtered. Since the structures present in the image become very coarse after several Gauss convolutions (Figure 1), the image size can be decreased to reduce memory usage and computational cost without a significant information loss. Every time the value of σ is doubled, a new octave in the scale space is created, meaning the image filtered using twice the initial value of σ is halved by omitting every second pixel (Figure 2). The newly created image is then used as the first image of the next octave of the scale space pyramid. The entire resizing process does not require any extra interpolation since all pixels in a Gaussian filtered image already contain information about their neighbouring pixels.

The difference of each pair of images separated by a multiplicative factor k in each scale space octave is then used to create Difference of Gaussians images (Figure 2)

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \tag{3}$$

These Difference of Gaussians (DoG) images provide a close approximation to images created by the Laplacian of Gaussian function. Later, the Gauss filtered images are used to form the feature descriptors efficiently.

To search for extrema in s images, $s+3$ images per octave have to be created in the initial Gauss pyramid. Lowe determined $s=3$ to be the most reasonable value for s. To generate s intervals per octave, the factor k has to be calculated using

$$k = 2^{1/s} \tag{4}$$

So, the images are convolved with $\sigma, 2^{1/s}\sigma, 2^{2/s}\sigma$ and so forth. Since the convolution mask should always be as big as 3σ, the calculations for each convolution would become more complex with every scale. However, instead of filtering the input image with different filter masks, Lowe showed that the images can also be filtered recursively. This means, every image in the image pyramid is created by filtering the previous image using the same σ. This leads to the same images being produced with a reduced process complexity.

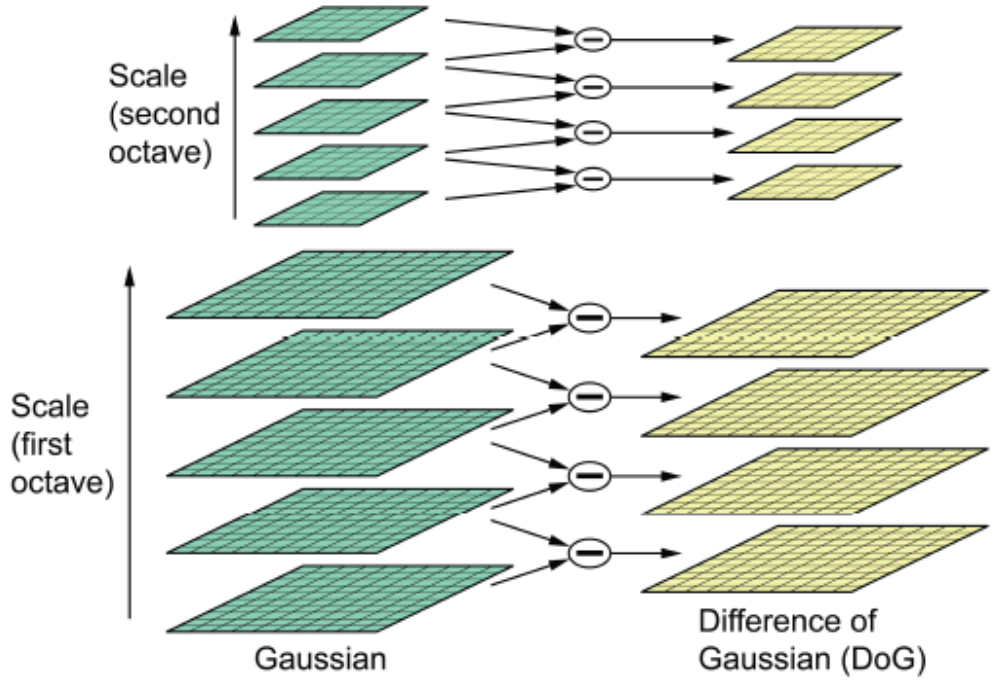


Figure 2: Generation of the Difference of Gaussians pyramid (Lowe, 2004).

Extrema Detection

Once a DoG pyramid has been created, extrema are searched in the images of each octave. To do this, each pixel of an image is compared to the eight pixels bordering it at the same scale and to the nine pixels in the scale above and below (Figure 3). Only if the pixel value is smaller or bigger than all of the 26 other pixels, it is accepted as an extremum. Since the first and the last image of each octave cannot be compared to all 26 pixels, no extrema are searched in those scales.

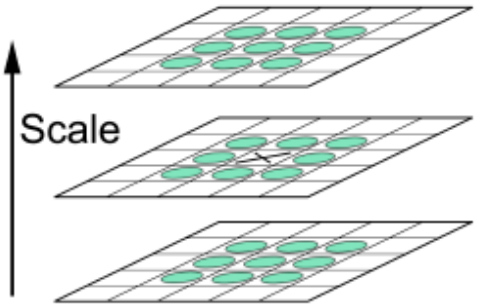


Figure 3: Extrema detection in scale space (Lowe. 2004).

2.2 Feature Localization

Once the extrema detection has been completed, unstable points need to be detected and filtered out. A straight forward approach is to use a normalized contrast threshold; a value of 0.03 has been suggested by Lowe. Considering the higher recognition value of corner points over edge points, corner points are preferred. To remove the edge points, the Hessian matrix is employed for all the detected points

$$H = \begin{pmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{pmatrix} \tag{5}$$

The derivatives D_{xx}, D_{yy} and D_{xy} are determined by calculating the differences between neighbouring points

$$D_{xx} = D(x+1, y) + D(x-1, y) - 2D(x, y) \tag{6}$$

$$D_{yy} = D(x, y+1) + D(x, y-1) - 2D(x, y) \tag{7}$$

$$D_{xy} = (D(x+1, y+1) - D(x-1, y+1) - D(x+1, y-1) + D(x-1, y-1))/4 \tag{8}$$

The trace

$$T_H = D_{xx} + D_{yy} \tag{9}$$

And the determinant

$$Det_H = D_{xx}.D_{yy} - (D_{xy})^2 \tag{10}$$

are then used to calculate the ratio of principal curvatures. This ratio is checked against a threshold r by using the following equation:

$$\frac{(T_H)^2}{Det_H} < \frac{(r+1)^2}{r} \tag{11}$$

Lowe uses $r = 10$ to reject points with a lower ratio of principal curvatures. Practically, this means that most points lying along straight edges are removed, while points lying on corners are accepted.

2.3 Sub-Pixel Localization

Lowe incorporates a method developed by M. Brown (Brown and Lowe, 2002) to interpolate the position of the feature in scale space. To achieve sub-pixel localization a Taylor expansion of the scale-space function $D(x; y; \sigma)$ is built up around the feature location

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x \tag{12}$$

This function is differentiated and set to zero to find the new, interpolated location Δx of the feature:

$$\Delta x = -\frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x} \tag{13}$$

in practice, the Hessian H and the derivative d used in this equation are approximated using pixel differences:

$$H = \begin{pmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{pmatrix} \tag{14}$$

where the Hessian matrix is calculated in the following ways

$$h_{00} = D_{n-1}(x, y) - 2D_n(x, y) + D_{n+1}(x, y)$$

$$h_{11} = D_n(x, y-1) - 2D_n(x, y) + D_n(x, y+1)$$

$$h_{22} = D_n(x-1, y) - 2D_n(x, y) + D_n(x+1, y)$$

$$h_{01} = h_{10} = \frac{1}{4}\left(D_{n+1}(x, y+1) - D_{n+1}(x, y-1) + D_{n-1}(x, y+1) + D_{n-1}(x, y-1)\right)$$

$$h_{02} = h_{20} = \frac{1}{4}\left(D_{n+1}(x+1, y) - D_{n+1}(x-1, y) + D_{n-1}(x+1, y) + D_{n-1}(x-1, y)\right)$$

$$h_{12} = h_{21} = \frac{1}{4}\left(D_n\left(x+1, y+1\right) - D_n\left(x-1, y+1\right) + D_n\left(x+1, y-1\right) + D_n\left(x-1, y-1\right)\right)$$

$$d = \begin{pmatrix} D_{n+1}\left(x, y\right) - D_{n-1}\left(x, y\right) \\ D_n\left(x, y+1\right) - D_n\left(x, y-1\right) \\ D_n\left(x+1, y\right) - D_n\left(x-1, y\right) \end{pmatrix} \tag{15}$$

The new interpolated position of the feature is obtained by solving the matrix equation $Hx = d$. The interpolated position change can vary in x and y direction on the image and in z direction within the image pyramid. This feature interpolation is repeated until the generated offset is smaller than 0.5 in any dimension (x, y and z). It also might be the case that the interpolation function does not converge for a specific feature, in that case the feature is discarded.

2.4 Orientation Assignment

To achieve rotation invariance, the gradient magnitude $m\left(x; y\right)$ and orientation $\theta\left(x; y\right)$ is computed for all scales $L\left(x; y\right)$ of the Gauss pyramid.

$$m\left(x, y\right) = \sqrt{\left(L\left(x+1, y\right) - L\left(x-1, y\right)\right)^2 + \left(L\left(x, y+1\right) - L\left(x, y-1\right)\right)^2} \tag{16}$$

$$\theta\left(x, y\right) = \tan^{-1}\frac{\left(L\left(x, y+1\right) - L\left(x, y-1\right)\right)}{\left(L\left(x+1, y\right) - L\left(x-1, y\right)\right)} \tag{17}$$

To find the dominant orientations a histogram with a bin size of 36 is created. The gradient orientations surrounding the feature, weighted by their gradient magnitude and a Gaussian window, centred at the feature location, are used to fill the histogram. Once the histogram has been created, the highest peak and all peaks bigger than 80% of the highest peak are selected as dominant orientations. Even though only about 15% of the points have more than one orientation, according to Lowe these additional orientations improve the stability of matching significantly. Finally, to achieve a higher accuracy, the positions of the peaks within the histogram are interpolated using a parabola approximated to the 3 closest values.

2.5 Local Image Descriptor

Having found the location, scale and rotation, a scale and rotation invariant descriptor of the feature can be created. Therefore a method based on complex neurons in primary visual cortex (Edelman et al., 1997) has been incorporated in the SIFT operator. The main benefit of complex neurons is their ability to respond even to slightly shifted gradients on the retina. This ability has been reproduced to create descriptors invariant to small local shifts. The gradient magnitudes and coordinates, calculated during the orientation assignment, are rotated around the feature relative to its found orientation. By rotating the area around the feature, rotation invariance is achieved. Scale invariance is assured by using the image gradients calculated at the same scale the feature belongs to. After that, the gradient magnitudes surrounding the feature are weighted by a Gaussian window to weaken the influence of gradients far away from the feature. In addition to that, the descriptor is not changed that much by a small shift, since gradients close to the center have a bigger impact. The σ of the Gaussian filter is set to half the size of the descriptor window, which itself has a size of 16*16. The descriptor window is divided up into 4*4 sub-windows to calculate 16 orientation histograms, each consisting of 8 bins (Figure 4). A trilinear interpolation is used to distribute gradients into bordering histogram bins, to once again reduce the effect of small local shifts on the descriptor. The final descriptor is a vector holding 128 elements (16 histograms composed of 8 bins). A 128 element vector has been found optimal according to experiments done by Lowe (2004).

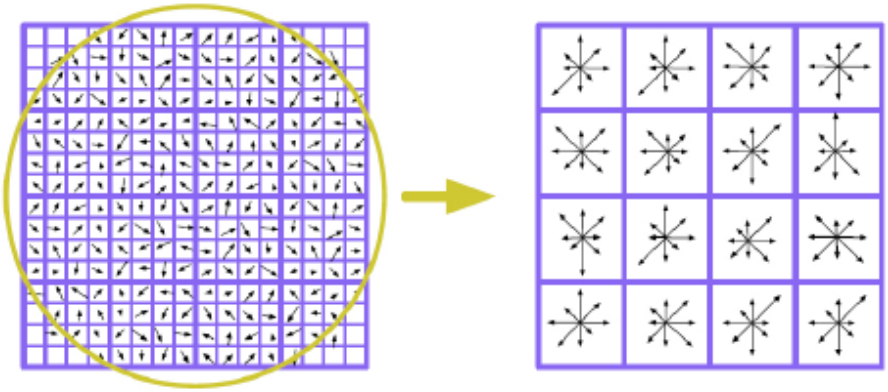


Figure 4: Sixteen histograms of the gradients of an area of 16x16 pixels. Each histogram consists of eight bins (symbolized by the arrows). The Gaussian window used to weight the histogram entries is depicted by the circle on the left (Zimmermann, 2007).

Processing till this step is sufficient to provide a rotation and scale invariant descriptor. To make the descriptor more robust to linear brightness changes, the elements of the vector are normalized to unit length. The influence of non-linear brightness changes (e.g. illumination changes that affect some surfaces more than others) is reduced by thresholding the vector elements. Lowe determined 0.2 to be a reasonable threshold. After that, all the values are normalized once again to unit length. By normalizing again, the distribution of orientations becomes more important for the descriptor than the change in relative magnitudes between some gradients.

2.6 Feature Matching

Even though the purpose of the SIFT operator is only the detection of features, Lowe also proposes a matching strategy for the found features in his paper (1999). To compare two descriptors with each other, the Euclidean distance d of the descriptor vectors x and y is calculated.

$$d = \sqrt{\sum_{i=1}^{128} (x_i - y_i)^2} \tag{18}$$

However it is not sufficient to simply find the matching point with the lowest Euclidean distance and accept all matches with a distance lower than a given threshold. The Euclidean distance between correct matches can vary a lot, so matching this way would mean that a lot of correct points would be omitted, while at the same time the number of incorrect matches would be increased. To compensate for this effect a more sophisticated matching strategy is utilized. Using descriptor distance as the criteria, for every feature the two closest matches in the other image are determined. If the distance of the second-closest match is smaller than 0.8 times the distance of the closest match, the first closest match is accepted as a legitimate match for a feature in the reference image. This mechanism works with high reliability, as it can be assumed that the distance between two incorrect matches is much shorter than the distance between a correct match and an incorrect match. According to Lowe, for optical images, this way 90% of the false matches are removed, while only 5% of actually correct matches are wrongly classified as bad detections. Since comparing the distance of all features with each other would not be efficient, Lowe uses an approximate algorithm called Best Bin First (BBF) (Beis and Lowe, 1997). BBF is

based on a modified version of the k-dimensional tree, where instead of searching all branches, by using a backtracking branch-and-bound search, only the bins lying close to the query point are searched. By only checking the 200 closest neighbours in the tree, the search speeds up drastically (2 orders of magnitude faster than exact nearest neighbour search for a database of 100.000 features) (Lowe, 1999). An example of two images matched using the scale and rotation invariant SIFT operator can be seen in Figure 5.

Figure 5: Matches fund in two test optical images using the SIFT operator (Schwind, 2008).

Bibliography

Beis, J., and Lowe, D.G., 1997. Shape indexing using approximate nearest neighbour search in high dimensional spaces, *in Proc. IEEE Conference on Computer Vision Pattern Recognition*, pp. 1000-1006, June 1997

Edelman, S., Intrator, N., and Poggio, T., 1997. Complex cells and object recognition. Unpublished manuscript available at http://www.cs.tau.ac.il/~nin/

Lowe, D.G., 1999. Object recognition from local scale-invariant features, *in Proc. International Conference on Computer Vision*, pp. 1150-1157, Corfu, Greece, Sep 1999

Lowe, D.G., 2004. Distinctive image features from scale-invariant features, *International Journal of Computer Vision*, 60(2), pp. 91-110

Mikolajczyk, K. and Schmid, C., 2004. Scale and affine invariant interest point detectors, *International Journal of Computer Vision*, 60, pp. 63-86

Schwind, P., 2008. Critical evaluation of the SIFT operator for remote sensing imagery, *Diplomarbeit*, Remote Sensing Technology Institute (IMF), German Aerospace Centre (DLR)

Zimmermann, F., 2007. Feature-detektion und deskriptoren-berechnung auf der grafikkarte. *Master's Thesis*, TU Kaiserslautern (in German)